

Online Budgeted Allocation with General Budgets

NATHANIEL KELL, Duke University
DEBMALYA PANIGRAHI, Duke University

We study the online budgeted allocation (also called ADWORDS) problem, where a set of impressions arriving online are allocated to a set of budget-constrained advertisers to maximize revenue. Motivated by connections to Internet advertising, several variants of this problem have been studied since the seminal work of Mehta, Saberi, Vazirani, and Vazirani (FOCS 2005). However, this entire body of work focuses on a single budget for every advertising campaign, whereas in order to fully represent the actual agenda of an advertiser, an advertising budget should be expressible over multiple tiers of user-attribute granularity. A simple example is an advertising campaign that is constrained by an overall budget but is also accompanied by a set of sub-budgets for each target demographic. In such a contract scheme, an advertiser can specify their true user-targeting goals, allowing the publisher to fulfill them through relevant allocations.

In this paper, we give a complete characterization of the ADWORDS problem for general advertising budgets. In the most general setting, we show that, unlike in the single-budget ADWORDS problem, obtaining a constant competitive ratio is impossible and give asymptotically tight upper and lower bounds. However for our main result, we observe that in many real-world scenarios (as in the above example), multi-tier budgets have a *laminar* structure, since most relevant consumer or product classifications are hierarchical. For laminar budgets, we obtain a competitive ratio of $e/(e-1)$ in the small bids case, which matches the best known ADWORDS result for single budgets. Our algorithm has a primal-dual structure and generalizes the primal-dual analysis for single-budget ADWORDS first given by Buchbinder, Jain, and Naor (ESA 2007). However many new ideas are required to overcome the barriers introduced by laminar budgets—our algorithm uses a novel formulation that overcomes non-monotonicity in the syntactically defined dual variables, as well as a dynamically maintained labeling scheme that tracks the “most-limiting” budgets in the hierarchy.

CCS Concepts: • **Theory of computation** → **Online algorithms**;

Additional Key Words and Phrases: Internet advertising, adwords, user targeting, primal-dual algorithms

1. INTRODUCTION

The online budgeted allocation problem, also called the ADWORDS problem, has had a significant impact on the theory and practice of online (Internet) advertising. In this problem, an advertisement publisher is tasked with matching user-generated advertisement slots on a web page (typically called *impressions*) to advertisers (typically called *bidders*). More formally, the publisher is given a set of offline bidders $u \in U$ at the outset of the problem, each of whom is specified by a budget B_u indicating the maximum revenue the publisher can receive from bidder u . A set of impressions $v \in V$ then arrive in an online sequence and must each be irrevocably assigned to a unique bidder. Upon assigning impression v to bidder u , the publisher receives revenue r_{uv}

This work was supported in part by NSF under contracts CCF 1527084 and AitF 1535972, a Google Faculty Research Award, and a Yahoo FREP Award.

Author's addresses: Department of Computer Science, Duke University; email: kell@cs.duke.edu, debmalya@cs.duke.edu.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

EC'16, July 24–28, 2016, Maastricht, The Netherlands.

ACM 978-1-4503-3936-0/16/07 ...\$15.00.

Copyright is held by the owner/author(s). Publication rights licensed to ACM.

<http://dx.doi.org/10.1145/2940716.2940770>

(typically called a *bid value*) from bidder u provided she has not reached her budget. The objective is to maximize the total revenue generated over all impressions.

The ADWORDS problem was introduced in the seminal work of Mehta, Saberi, Vazirani, and Vazirani [Mehta et al. 2007]. While this problem generalizes the classic *online matching* problem introduced by Karp, Vazirani, and Vazirani [Karp et al. 1990], research has focused on the so called *small bids* case, i.e., on algorithmic performance as the ratio $\frac{\max_{v \in V} b_{uv}}{B_u}$ tends to 0 for all advertisers. This assumption models most real-world scenarios, where the revenue generated from a single impression is infinitesimal compared to the total budget of an advertiser. The small bids assumption distinguishes ADWORDS from online matching and makes them incomparable from a technical perspective; given its natural applicability and popularity in Internet advertising, we will also focus primarily on the small bids case.

Since the introduction of these problems, several variants of ADWORDS and online matching have been studied, motivated primarily by the evolving challenges advertisement publishers face in practice. For example, Agrawal and Devanur [Agrawal and Devanur 2015] recently considered arbitrary linear and non-linear convex budget constraints for the stochastic input model, and Devanur and Jain [Devanur and Jain 2012] studied concave returns on revenue, both motivated by problem features such as under-delivery penalties and pay-per-click advertisements. Another motivation for these variants that is receiving increasing attention is that of *impression diversification* or *representativeness*. More specifically, a campaign contract is usually of the form: “deliver ten million advertisements to Californian females in the month of July.” Although such an agreement clearly indicates a target group to which the publisher should restrict its assignments, often the advertiser still wants the impressions to be equally spread among the sub-populations of the targeted group (e.g., in the above contract, the advertiser will likely be unhappy if their ads are only shown to white females in their twenties living in Los Angeles). Although it can be at odds with short-term revenue gains, ensuring diversity is crucial with respect to *long-term* revenue for the publisher, as advertisers that see a high return on investment (in this case via reaching desired audiences) are more likely to continually purchase future contracts.

There are several recent works that address impression diversification (e.g., see [Ghosh et al. 2009], [Bharadwaj et al. 2012], and [Hojjat et al. 2014]). In many of these results, the objective function of the problem incorporates a diversity penalty, usually in the form of a distance function that incurs a cost if the algorithm’s assignment differs significantly from the advertiser’s ideal allocation. However another natural approach, which to the best of our knowledge has yet to be considered, is a scheme where contracts are specified over multiple tiers of user-attribute granularity. Recalling our previous contract example, in a multi-tier scheme the advertiser could further specify constraints in terms of age groups, e.g., “no more than four million of the total ten million advertisements should be shown to age groups 20-29, 30-39, and 40+, respectively.” Further constraints could also be placed on each age group in terms of residency, e.g., “of the maximum four million advertisements assigned in each age group, no more than two million should be shown to residents in Los Angeles.” This scheme has several appealing features, the foremost being that it allows advertisers to explicitly indicate their true user-targeting goals with a high degree of expressibility (whereas penalty functions often assume that the ideal allocation must follow the same distribution as the overall targeted population, i.e., is forced to be a representative sample).

With this motivation, we introduce a generalization of the ADWORDS problem that we call ADWORDS *with general budgets* (denoted ADGENERAL). As earlier, there is an offline set of advertisers U and a set of impressions V that arrive online. We also

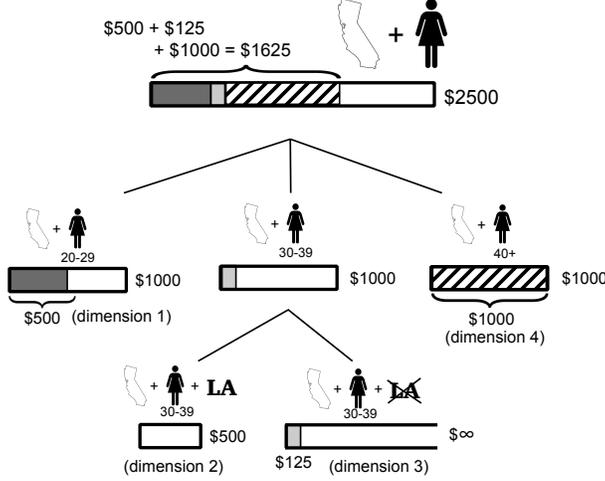


Fig. 1. Illustration of how the “Californian female” example translates into a bidder’s budget in an instance of ADGENERAL (which is also an ADLAMINAR instance). For simplicity, we only have a Los Angeles residency subdivision for the 30-39 age range, giving us four dimensions in total. Observe that $K_s = \{2, 3\}$ for the 30-39 age-range budget, and $K_s = \{1, 2, 3, 4\}$ for the overall budget; furthermore, we have that $B_u^{\{2,3\}} = 1000$ and $B_u^{\{1,2,3,4\}} = 2500$. The algorithm has currently earned \$500, \$125, and \$1000 on dimensions 1, 3, and 4, respectively (and \$0 on dimensions 2); therefore, we have used \$1625 out of the overall budget’s capacity of \$2500. If an impression v is newly assigned to this bidder u such that $r_{uv}^{(1)} = \$1$, $r_{uv}^{(4)} = \$1$, and $r_{uv}^{(k)} = 0$ for $k = 2, 3$, then after the assignment we will have earned \$501 on dimension 1 and thus \$1626 overall (note that we do not earn the \$1 of revenue from dimension 4 since its budget is at capacity).

have a set of *dimensions* K_u for each bidder u that represent the smallest level of user-attribute granularity over which the bidder defines her budgets. In our previous example, one dimension would correspond to “Californian females living in Los Angeles in the age range 20-29”; another dimension would be “Californian females living outside of Los Angeles in age range 40+”. For simplicity of notation, we will consider a universal set of dimensions $K = \cup_u K_u$ and assume that each bidder has this same set of dimensions; dimensions in $K \setminus K_u$ will simply earn no revenue and have no budget constraints for bidder u . If an impression v is assigned to an advertiser u , then the algorithm earns $r_{uv}^{(k)}$ (called the *bid value* in dimension k) on dimension k for every $k \in K$. The revenue generated from an advertiser u is subject to a set of budget constraints S_u , where each constraint $s \in S_u$ caps the total revenue generated from a subset of dimensions K_s to a budget $B_u^{(s)}$ (see Figure 1 to see how subsets of dimensions are used to define the budgets in our example). As usual, the objective in the ADGENERAL problem is to maximize the total revenue generated by the algorithm. We will measure our algorithms using *competitive analysis*, which is the maximum ratio over all instances between the objective of an optimal solution and the algorithmic solution (see, e.g., [Borodin and El-Yaniv 1998]).

We point out that we allow multiple non-zero $r_{uv}^{(k)}$ revenue entries for a single impression-bidder pair (u, v) . This is to reflect the fact that certain categories are more definite than others. For instance, the campaign that an advertisement belongs to is known but attributes of an online user may be less certain. If an anonymous online user has a 60% chance of being below 30 and 40% of being above 30 (e.g., based on

browsing behavior), then the revenue earned for this user should be split in the same proportion between these two user category dimensions. In this case, the bid vector has a non-zero entry in multiple dimensions.¹ Also observe that for an assignment with multiple non-zero dimensions, it is possible that the algorithm might only earn revenue on some dimensions and not others. More specifically, if an assignment is made such that a non-zero dimension k belongs to a budget that is currently at capacity, the algorithm *does not* earn the revenue from dimension k (this is illustrated by Figure 1). This is a natural way of defining revenue derivation in a setting where demographics are uncertain. For example, following the setup in Figure 1, suppose N female users arrive in the system where each has a 50% chance being in the 20-29 age range and 50% chance of being in the 40+ age range (so in expectation there are $N/2$ users in each category). If the 40+ budget is currently at capacity and the bid value for these users is \$2, then the algorithm earns \$1 on the 20-29 dimension and \$0 for the 40+ dimension for all N impressions (provided the 20-29 budget never reaches capacity). Note that this is exactly what the advertiser was willing to pay this group of users: N dollars for the expected $N/2$ users in the 20-29 age range and \$0 for the remaining 40+ users.

The first natural question to consider is whether this generalization changes the structure of the ADWORDS problem. To understand this, let us consider an instance with a single advertiser. In this case, an algorithm that assigns all impressions to the lone advertiser is clearly optimal for the ADWORDS problem. However, let us now consider the ADGENERAL problem with 3 dimensions, and budget constraints of \$1 each on dimensions $\{1, 2\}$ and $\{2, 3\}$. Now, suppose the first impression has a revenue of \$1 on dimension 2 alone. Should the algorithm assign the impression to the lone advertiser? If it does, then the instance will generate two impressions yielding a revenue of \$1 on dimensions 1 and 3 each, while if it does not, then the instance will generate no other impression. Clearly, this example shows that no algorithm can do better than a competitive ratio of 2, even with a single advertiser. One may object that the small bids assumption is being violated, but replacing an impression of bid value \$1 with $1/\epsilon$ impressions of bid value $\epsilon \rightarrow 0$ still produces a constant (deterministic) lower bound of $3/2$. In fact, this lower bound is a manifestation of a more general observation: *it may be a better option for the algorithm to not allocate an impression, or to not earn revenue from some of the dimensions, even when possible to do so.* This is in sharp contrast to the classical ADWORDS problem, where an impression should always be allocated if possible. Thus, the ADGENERAL problem introduces an aspect of “admission control” to the ADWORDS framework. Our results for ADGENERAL are characterized in the following theorem, which is shown in Section 3.

THEOREM 1.1. *The competitive ratio of the ADGENERAL problem is $\Theta(\lg p)$ under the small bids assumption, where $p = \max_{u \in U, k \in K} |\{s \in S_u : k \in K_s\}|$ denotes the maximum number of budget constraints to which any dimension belongs.*

Although there is a super-constant lower bound for ADGENERAL, we observe that many multi-tier budgets based on real-world instances will have additional structure. As we saw earlier in our motivating example, each budget constraint was a subdivision of a more general constraint, i.e., the total budget for Californian females was

¹The reader is referred to Chatwin [Chatwin 2013] for an overview on uncertainty in determining online user attributes. Adhikari and Dutta discuss how attribute uncertainty is weighted in real-time bidding strategies [Adhikari and Dutta 2015], while Ghosh *et al.* [Ghosh et al. 2009] highlights the different levels of information about the users available to different entities in Internet advertising. Techniques for determining user attributes based on historical and prior behavior have also been extensively studied in the marketing research community (see Barajas Zamora [Barajas Zamora 2015] and references contained therein), where the mapping of user behavior to attributes is inherently probabilistic.

divided into age groups to obtain the second level of budgets, and then each of these age groups was divided based on residency to obtain the next level. In essence, many consumer and product classifications are naturally hierarchical. If the budgets of an ADGENERAL instance are defined over such a taxonomy, then we have the additional structure that the budget sets $\{K_s : s \in S_u\}$ for each bidder u form a *laminar* family, i.e., for every pair of intersecting sets in $\{K_s : s \in S_u\}$, one is contained in the other.

Thus, we also consider the ADWORDS *with laminar budgets* problem (ADLAMINAR). It turns out that laminar budgets make admission control redundant — the algorithm can now earn revenue whenever possible. However, there are other conceptual difficulties. Consider an instance with 2 dimensions, where an advertiser has a budget of \$1 for dimension 1 and an overall budget of \$2 for dimensions $\{1, 2\}$. At any point in the algorithm, what is the total budget of dimension 2, i.e., the capacity on the total revenue earned from dimension 2? This value clearly depends on the revenue earned on dimension 1, and therefore changes during the course of the algorithm. This is in sharp contrast to the classical ADWORDS setting where the total budget of a bidder remains unchanged (note the distinction between total budget and remaining budget). The first technical hurdle, therefore, is to define a dynamic notion of total budget on individual dimensions. In addition, we also need to define a notion of current budget utilization for individual dimensions to determine which dimensions we should prefer in making the allocation. Again, this notion is canonical in the classical ADWORDS problem — it is simply the fraction of a budget that has already been earned as revenue. In our more general setting, a single dimension might be in multiple budget constraints, and therefore we must first identify the most constraining budget. Once we do so, we need a mechanism for importing the budget utilization of this constraining budget to the dimension itself. For instance, in the example above, if the revenue earned on dimension 1 is \$1 and that on dimension 2 is \$0 at some point in the algorithm, then should the budget utilization for dimension 2 be 0.5 (from its most constraining budget constraint) or 0 (from the fact that the algorithm has not earned any revenue at all from dimension 2 yet)? It turns out that these concepts (revenue capacity, most constraining budget, and the budget utilization of a dimension) are closely tied to each other and have to be defined through a common inductive process. Our main technical contribution for the ADLAMINAR problem is to carefully define these entities in a way that ensures semantic consistency and eventually gives our main result for this problem: an algorithm with a competitive ratio of $e/(e - 1)$, matching that for the classic ADWORDS problem [Mehta et al. 2007]. This result is proved in Section ??.

THEOREM 1.2. *The competitive ratio of ADLAMINAR is $e/(e - 1)$ under the small bids assumption.*

Finally, we study the ADGENERAL problem *without* the small bids assumption. In the absence of this assumption, there are two possible variants — either (a) the algorithm can choose the amount of revenue it earns on any given dimension (which can be less than the corresponding bid value) from an impression, or (b) the algorithm is constrained to earn the entire bid value as revenue on any dimension, which means that an assignment of an impression to an advertiser is only allowed if adding the bid value to the previously earned revenue on each dimension does not violate any constraint. The former is more natural in the context of Internet advertising — we call it the ADGEN-P problem² and match the bounds in the small bids case.

²P for “partial”

THEOREM 1.3. *The competitive ratio of the ADGEN-P problem is $\Theta(\lg p)$, where $p = \max_{u \in U, k \in K} |\{s \in S_u : k \in K_s\}|$ denotes the the maximum number of budget constraints to which any dimension belongs.*

We also study the latter problem (where the entire bid value is always added to the revenue), primarily because of interesting connections to the classical online admission control problem [Awerbuch et al. 1993]. In Theorem 1.4, we give the competitive ratio of the ADGEN-AON problem (AON for “all or nothing”). As a byproduct of our result, we also obtain tight bounds for the online admission control problem, slightly improving the classical bounds of [Awerbuch et al. 1993].

THEOREM 1.4. *Let $\epsilon = \max_{u,v,s} \frac{\sum_{k \in K_s} r_{uv}^{(k)}}{B_u^{(s)}}$ denote the maximum bid-to-budget ratio and $p = \max_{u \in U, k \in K} |\{s \in S_u : k \in K_s\}|$ denote the maximum number of budget constraints to which any dimension belongs. If $\frac{1}{\lg(2p)} < \epsilon < 1$, then the competitive ratio of the ADGEN-AON problem is $\Theta\left(\frac{p^{\frac{\epsilon}{1-\epsilon}}}{\epsilon}\right)$.*

Note that we do not consider $\epsilon \leq \frac{1}{\lg(2p)}$ in Theorem 1.4, as in this case ADGEN-AON is essentially identical to ADGENERAL with small bids.

For reasons of brevity, we do not discuss the admission control problem – the implication of the above theorem to this problem is straightforward – and relegate the details for the large bids case, both ADGEN-P and ADGEN-AON, to the full version of the paper.

Related Work. Given the large volume of work in this area, we will only sample a small fraction of the online matching and ADWORDS literature, focusing on results in the (adversarial) online model. For a comprehensive survey, including results in stochastic input models, the reader is referred to the survey by Mehta [Mehta 2013].

Karp, Vazirani, and Vazirani [Karp et al. 1990] introduced the online matching problem and gave a tight $e/(e-1)$ -competitive algorithm (see also [Goel and Mehta 2008], [Birnbbaum and Mathieu 2008], and [Devanur et al. 2013]). The first generalization was to the b -matching problem by Kalyanasundaram and Pruhs [Kalyanasundaram and Pruhs 2000]. Later generalizations include a vertex-weighted version by Aggarwal *et al.* [Aggarwal et al. 2011], a *pay-per-click* model using stochastic rewards by Mehta and Panigrahi [Mehta and Panigrahi 2012] (see also [Mehta et al. 2015]), and a bi-objective model suggested by Aggarwal *et al.* [Aggarwal et al. 2014]. Devanur and Jain [Devanur and Jain 2012] explored non-linear concave objectives to encode, e.g., penalties for under-delivery. In terms of techniques, most of the initial results used combinatorial methods, but recent work has focused on a (randomized) primal dual technique introduced by Devanur *et al.* [Devanur et al. 2013].

The ADWORDS problem, which generalizes b -matching, was introduced by Mehta *et al.* [Mehta et al. 2007], who gave an $e/(e-1)$ approximation for small bids. They also showed that this competitive ratio is the best possible. Without the small bids assumption, the greedy algorithm for the ADWORDS problem has a competitive ratio of 2, and while this is tight for deterministic algorithms, obtaining a better ratio using a randomized algorithm is open. Buchbinder *et al.* [Buchbinder et al. 2007] gave an alternative primal-dual analysis for the algorithm of Mehta *et al.* [Mehta et al. 2007] with the same competitive ratio. More recently, other variants of the ADWORDS problem have been considered. For instance, Feldman *et al.* [Feldman et al. 2009] and Aggarwal *et al.* [Aggarwal et al. 2011] introduced variants to model display ads with vertex weights and/or capacities.

2. ADWORDS WITH LAMINAR BUDGET CONSTRAINTS (ADLAMINAR)

Recall the ADLAMINAR problem: we given a set of offline bidders U and a set of impressions V that arrive online, where each bidder-impression pair (u, v) is specified by a bid value $r_{uv}^{(k)}$ for each dimension $k \in K$. The revenue generated from a bidder u is subject to an arbitrary set of budget constraints S_u , where each constraint $s \in S_u$ caps the total revenue generated from a subset of dimensions K_s to a given budget $B_u^{(s)}$. We assume that the sets $\{K_s : s \in S_u\}$ form a laminar family, i.e., for every pair of intersecting sets in $\{K_s : s \in S_u\}$, one is contained in the other.

In this section, we give an algorithm for the ADLAMINAR problem with a competitive ratio of $e/(e-1)$ under the small bids assumption (Theorem 1.2). This bound is tight because of a matching lower bound for the ADWORDS problem [Mehta et al. 2007]. Throughout this section, we assume that for every dimension $k \in K$, a constraint s with $K_s = \{k\}$ appears in S_u for each bidder u . This is wlog since a budget can be made arbitrarily large. We will call these constraints *singleton budgets* of bidder u .

2.1. Primal and Dual Formulations

Our algorithm uses a primal-dual formulation of the ADLAMINAR problem. In other words, we give a primal LP formulation of the ADLAMINAR problem and its corresponding dual and update the solutions to both LPs on the arrival of a new impression. The primal updates, which are guided by the dual solution, define the algorithm. The main challenge is to show that the dual updates maintain feasibility while ensuring that the ratio of the primal and dual objectives remains bounded by the desired competitive ratio $\rho = e/(e-1)$.

Let us define the *current budget utilization* for constraint s of bidder u (denoted $\kappa_u^{(s)}$) to be the fraction of budget $B_u^{(s)}$ currently used by the algorithm, or formally, $\kappa_u^{(s)} = \sum_{k \in K_s, v} r_{uv}^{(k)} / B_u^{(s)}$ for impressions v assigned to bidder u thus far. Let us call a dimension k *active* for bidder u if for all budgets $s \in S_u$ such that $k \in K_s$, the algorithm currently has $\kappa_u^{(s)} < 1$. In other words, u 's active dimensions are the ones on which the algorithm can still receive revenue from u .

As pointed out in the introduction, an algorithm for ADLAMINAR might gain revenue from only a subset of dimensions when assigning an impression to a bidder. To implement this flexibility in the LP, we introduce the notion of *assignment types*. Let $t \subseteq K$. For impression v and bidder u , we define a *type- t assignment* as one where the dimensions in t are active and the dimensions in $K \setminus t$ are inactive. Thus, our decision variables for the LP will be of the form x_{uvt} , where the algorithm sets x_{uvt} to be 1 if impression v is assigned to bidder u using a type- t assignment (and 0 otherwise). We then define $r_{uvt}^{(k)} = r_{uv}^{(k)}$ if $k \in t$; otherwise, $r_{uvt}^{(k)} = 0$. Our primal LP P is now defined as:

$$\begin{aligned} \max \quad & \sum_{u,v,t} x_{uvt} \sum_k r_{uvt}^{(k)} \\ \forall u \in U, s \in S_u : \quad & \sum_{v,t} x_{uvt} \sum_{k \in K_s} r_{uvt}^{(k)} \leq B_u^{(s)} \end{aligned} \tag{1}$$

$$\forall v \in V : \quad \sum_{u,t} x_{uvt} \leq 1 \tag{2}$$

$$\forall u \in U, v \in V, t \subseteq K : \quad x_{uvt} \geq 0.$$

Note that Eq. (10) ensures that the algorithm receives no revenue from inactive dimensions for a bidder, and Eq. (11) ensures that every impression v is assigned using a single type to a single bidder.

The dual D of this LP is defined as:

$$\begin{aligned}
& \min \sum_u \sum_{s \in S_u} \alpha_u^{(s)} B_u^{(s)} + \sum_v \sigma_v \\
\forall u \in U, v \in V, t \subseteq K : & \quad \sum_{s \in S_u} \left(\alpha_u^{(s)} \sum_{k \in K_s} r_{uvt}^{(k)} \right) + \sigma_v \geq \sum_k r_{uvt}^{(k)} \quad (3) \\
\forall u \in U, s \in S_u : & \quad \alpha_u^{(s)} \geq 0 \\
\forall v \in V : & \quad \sigma_v \geq 0.
\end{aligned}$$

Unfortunately, the dual stated above cannot be used directly in a primal dual algorithm. If a constraint s has budget utilization $\kappa_u^{(s)} = 1$, then the dual variable $\alpha_u^{(s)}$ also needs to be equal to 1 in order to balance the contributions to the two sides of the dual constraint by dimensions $k \in K_s$. (Note that the primal objective does not increase for these dimensions and hence the value of σ_v in the dual objective cannot depend on these dimensions either, if the ratio of the primal to dual objective is to be maintained.)

Now, if we naively enforce $\alpha_u^{(s)} = 1$ once a budget $B_u^{(s)}$ is tight, then the primal-dual ratio could be proportional to the number of nested levels, if a nested set of budgets are all tight. To obtain a constant competitive ratio, what our scheme will (roughly speaking) need to do is only set $\alpha_u^{(s)}$ to 1 at the highest level of nesting for each nested set of tight constraints. If we think of the primal objective as being “attributed” to dual variables in order to maintain the primal dual ratio, then what we want is that at any point of time, the primal objective from a given dimension for some bidder u is attributed to a unique dual variable representing a budget constraint for u containing that dimension. However, in order to implement this property in an online setting, $\alpha_u^{(s)}$ variables need to be non-monotone since the budgets in lower nesting levels might become tight first followed by the higher levels. Therefore, we need a means of raising and lowering each $\alpha_u^{(s)}$ so that at the end of the instance, the revenue earned from a dimension for bidder u is attributed to exactly one of these variables. In general, non-monotonicity of dual variables is undesirable in online algorithms because a satisfied dual constraint might become unsatisfied later. To overcome this problem, we give a new dual D' where we decompose $\alpha_u^{(s)}$ into decision variables that are indeed monotone in our eventual primal-dual analysis.

Formally, our transformed dual D' is defined as follows. Since each S_u is laminar, we can represent its set system as a forest F_u , where each node in the forest corresponds to a constraint $s \in S_u$, and the singleton budgets are the leaves. Let A_s be the set of ancestors of s in F_u including s itself. Define a new decision variable $\gamma_u^{(s)} = \sum_{s' \in A_s} \alpha_u^{(s')}$, and let $p(s)$ be the parent budget of s . Observe that $\alpha_u^{(s)} = \gamma_u^{(s)} - \gamma_u^{(p(s))}$ (where for a maximal set s with no parent in F_u , we set $\gamma_u^{(p(s))} = 0$). Using these new variables, we can rewrite Eqn. (12) in our original dual formulation as:

$$\forall u \in U, v \in V, t \subseteq K : \quad \sum_{s \in S_u} \left(\left(\gamma_u^{(s)} - \gamma_u^{(p(s))} \right) \sum_{k \in K_s} r_{uvt}^{(k)} \right) + \sigma_v \geq \sum_k r_{uvt}^{(k)}. \quad (4)$$

Next, we observe that the outermost summation on the LHS of Eqn. (13) telescopes, and the only remaining $\gamma_u^{(s)}$ are those for singleton budgets. This gives us our final dual formulation D' :

$$\begin{aligned} & \min \sum_v \sigma_v + \sum_u \sum_{s \in S_u} B_u^{(s)} \left(\gamma_u^{(s)} - \gamma_u^{p(s)} \right) \\ \forall u \in U, v \in V, t \subseteq K : & \sum_k \gamma_u^{(\{k\})} r_{uvt}^{(k)} + \sigma_v \geq \sum_k r_{uvt}^{(k)} \quad (5) \\ \forall u \in U, s \in S_u : & \gamma_u^{(s)} - \gamma_u^{(p(s))} \geq 0. \quad (6) \end{aligned}$$

2.2. Labeling Scheme

Recalling our above discussion, our goal will be to attribute the revenue earned on a dimension k for bidder u to exactly one dual variable $\gamma_u^{(s)}$, ideally to the $\gamma_u^{(s)}$ corresponding to the “most-limiting” budget $B_u^{(s)}$ such that $k \in K_s$. This suggests that we should make $\gamma_u^{(s)}$ a monotonically increasing function of the budget’s current utilization $\kappa_u^{(s)}$. However, simply using utilization to define $\gamma_u^{(s)}$ does not capture the interactions between budgets in the laminar setting. The overarching issue with just using $\kappa_u^{(s)}$ is the fact that $B_u^{(s)}$ might be the most utilized constraint for only *some* of the dimensions in K_s , since other budgets that sit below $B_u^{(s)}$ in the hierarchy may have higher utilization. This raises the following question: should the revenue currently constrained by these descendant budgets, say revenue earned on some particular dimension k' , affect how the algorithm determines the extent to which $B_u^{(s)}$ limits other unbounded dimensions like k ? The answer is not immediate. It is tempting to say “no” since the dimension- k' revenue is already bounded by a tighter budget; on the other hand, $B_u^{(s)}$ might in fact become the tightest budget for dimension k' later in the instance and ignoring the dimension- k' revenue till that time will prevent a smooth transition of the tightest budget for k' .

To overcome this challenge, we introduce a labeling scheme $\ell_u^{(s)} : \cup_u S_u \rightarrow [0, 1]$. Label $\ell_u^{(s)}$ for budget $B_u^{(s)}$ will represent the modified notion of the budget’s utilization that we need to properly measure the remaining capacity for future revenue. Our primal-dual analysis will then follow by making each dual variable a monotone function of these labels.

More concretely, we address the above challenge by having our labels maintain the following two high-level features:

- For label $\ell_u^{(s)}$ and bidder u , revenue from a dimension $k \in s$ will only contribute to the label if $\ell_u^{(s)}$ is at least as large as the labels of all budgets containing k that are subsets of s . This corresponds to identifying the “most constrained” budget for any dimension by interpreting these abstract labels as surrogates of actual budget utilizations.
- In defining label $\ell_u^{(s)}$, we need to identify the capacity of constraint s for future revenue earnings from the dimensions that are deriving their label from s . We define this capacity as the total budget $B_u^{(s)}$ minus the budgets of constraints below s that have a higher label. This automatically discounts the revenue earning capacities of dimensions that are deriving labels from descendant constraints of s .

One challenge with maintaining these properties is that they are somewhat circular. To determine the value of a label, we need to first determine which dimensions count toward the label, but determining dimension inclusion requires comparisons between label values. Another challenge is maintaining smoothness. As impressions are assigned to bidders, budgets that were previously slack will become tight, which re-

quires us to reassign dimensions to labels and change their capacities. In order to make our primal-dual analysis smooth, we will need to ensure that labels remain consistent after we reassign dimensions to labels and change label capacities.

To overcome the issue of circularity, we will not give an explicit label definition but rather give a set of label properties that we maintain throughout the algorithm. These properties are based on two sets of budgets, $L(s)$ and $T(s)$, that the algorithm will dynamically update for all bidders u and budgets $B_u^{(s)}$ (we drop the subscript u for simplicity). The two sets partition the descendant dimensions of s (i.e., each descendant dimension belongs to exactly one set in $L(s) \cup T(s)$). Intuitively, $L(s)$ contains singleton budgets $\{k\}$ representing dimensions that count toward label $\ell_u^{(s)}$. On the other hand, $T(s)$ contains the closest descendants of s that have a larger label than s , i.e., every dimension in s that is not in $L(s)$ derives its label from a budget in $T(s)$ or from one their respective descendants.

Let $R_u^{(k)}$ be the total revenue currently earned on dimension k for bidder u . Formalizing the above discussion, we say the labels for bidder u are *valid* if the following three properties hold for all $s \in S_u$.

- (1) *Property 1:* For all $\{k\} \in L(s)$, all constraints s' on the path from $\{k\}$ to s in F_u have $\ell_u^{(s')} \leq \ell_u^{(s)}$.
- (2) *Property 2:* For all $s' \in T(s)$, we have $\ell_u^{(s')} > \ell_u^{(s)}$, and for all s'' on the path from s' to s , we have $\ell_u^{(s'')} \leq \ell_u^{(s)}$.
- (3) *Property 3:* The following identity holds:

$$\ell_u^{(s)} = \frac{\sum_{\{k\} \in L(s)} R_u^{(k)}}{B_u^{(s)} - \sum_{s' \in T(s)} B_u^{(s')}}. \quad (7)$$

Note that once we have fixed sets $T(s)$ and $L(s)$ for all s , we can verify all three properties and use Eqn. (16) to directly compute each label for each $s \in S_u$. Also observe we can initialize all labels to be 0, and set $T(s) = \emptyset$ and $L(s) = \{\{k\} : k \in K\}$ to start with a valid labeling. Finally, we note that we will soon show that all labels remain non-negative (in the proof of Lemma 2.3).

Next, we define a procedure for updating labels after the revenue earned in a single dimension increases infinitesimally. More specifically, suppose an impression v is assigned to bidder u , and assume we have a valid labeling for the sets in S_u before the assignment. The assignment of the impression changes the values of $R_u^{(k)}$ for the active dimensions k . This necessitates label updates, which we define for an infinitesimally small increment in the value of $R_u^{(k)}$ for a particular dimension k . Note that the overall assignment of the impression is a sequence of such incremental changes.

The labels that we increase on such an increment are $\ell_u^{(s)}$ such that $\{k\} \in L(s)$, i.e., we will increase $R_u^{(k)}$ in the numerator of Eqn. (16) for all such $\ell_u^{(s)}$ and leave $T(s)$ and $L(s)$ fixed. If the labeling is valid before a given increment, and if after the increment the relative ordering of all $\ell_u^{(s)}$ remains the same, then by the definition of Properties 1 and 2 the labeling remains valid. Thus, in terms of updating $L(s)$ and $T(s)$, we only need to consider when the relative order of labels changes as a result of adding to $R_u^{(k)}$. In particular, there are two types of reordering that need to be considered. We will call these two reordering possibilities Events 1 and 2 and describe how the algorithm updates $L(s)$ and $T(s)$ in each case. Later, we will show that these set redefinitions maintain the current value of the label.

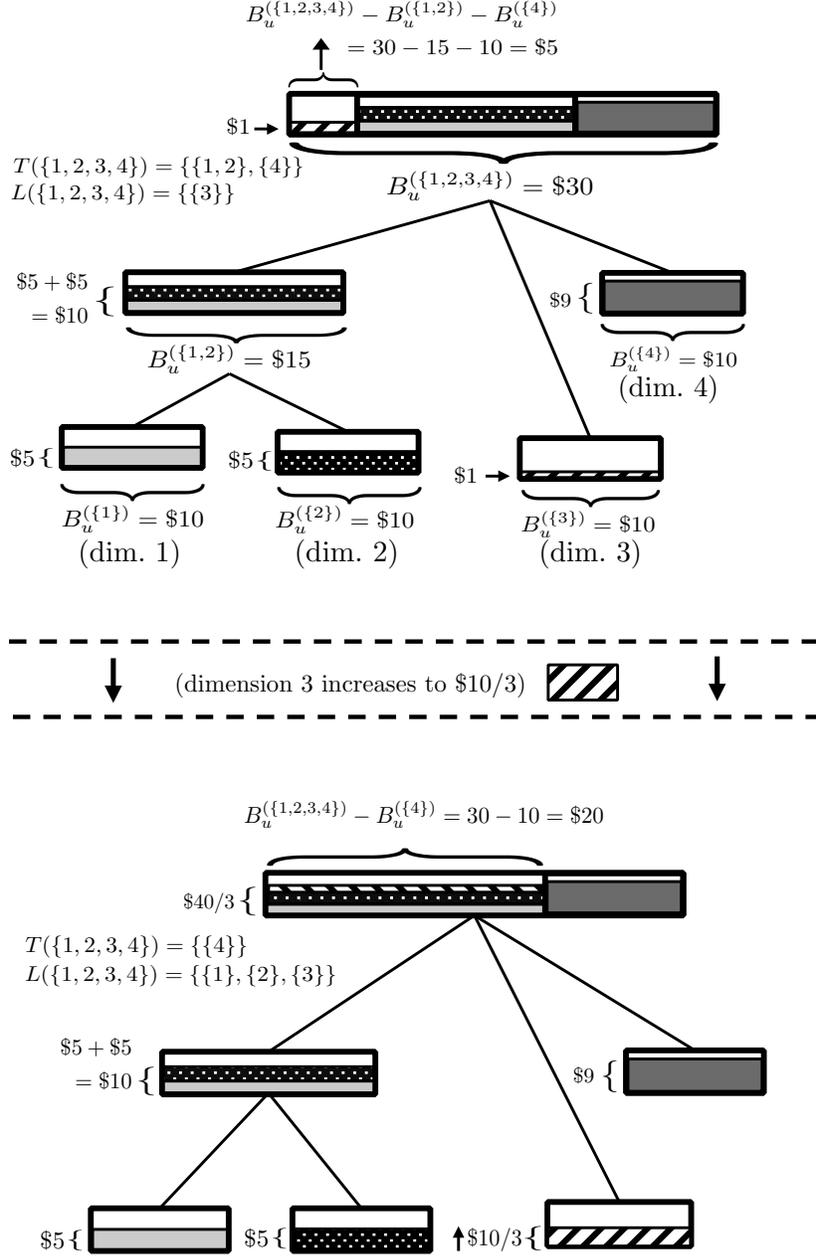


Fig. 2. Illustration of the modifications made to $L(s)$ and $T(s)$ in Event 2, occurring between budget sets $s = \{1, 2, 3, 4\}$ and $s' = \{1, 2\}$. In this example, the algorithm is incrementing the revenue on dimension 3. The state of the labels before dimension 3 revenue is added is shown on top. Notice that $\{3\} \in L(\{1, 2, 3, 4\})$ since $\ell_u^{\{3\}} = 1/10$ is smaller than $\ell_u^{\{1,2,3,4\}} = 1/5$ (note that the value of $\ell_u^{\{1,2,3,4\}}$ corresponds to the left most rectangle, where the middle and right rectangles show the revenue levels of the dimensions that are *not* included in the label). After increasing dimension 3 to $\$10/3$, we have that $\ell_u^{\{1,2,3,4\}} = 2/3$, which means it is about to surpass label $\ell_u^{\{1,2\}}$, therefore triggering Event 2. The state of the labels after the Event 2 modifications are shown on bottom (at the transition point). Notice that $\{1, 2\}$ has been removed from $T(\{1, 2, 3, 4\})$, and we have added the sets in $L(\{1, 2\})$ to $L(\{1, 2, 3, 4\})$ (namely, $\{1\}$ and $\{2\}$).

- **Event 1:** For some s such that $\{k\} \in L(s)$, there now exists a descendant s' of s such that $\ell_u^{(s')} > \ell_u^{(s)}$, where s' is on the path from $\{k\}$ to s and $\ell_u^{(s')} \leq \ell_u^{(s)}$ previously. In this event, s' is now added to $T(s)$. All singleton-budgets $\{k\}$ that are descendants of s' and belong to $L(s)$ are now removed from $L(s)$. Additionally, all descendants of s' that belong to $T(s)$ are also removed from $T(s)$.
- **Event 2:** For some constraint s such that $\{k\} \in L(s)$, there now exists a descendant s' of s such that $\ell_u^{(s)} \geq \ell_u^{(s')}$, where $s' \in T(s)$ previously. In this event, s' is removed from $T(s)$. Conversely to Event 1, all constraints in $T(s')$ are added to $T(s)$, and all $\{k\} \in L(s')$ are added to $L(s)$.

In order to make the process smooth, we will think of the updates in Events 1 and 2 as being done at the transition point where $\ell_u^{(s)} = \ell_u^{(s')}$. This completes the description of our labeling scheme and the process by which the algorithm determines them. We encourage the reader to refer to Figure 3 for a small example of an Event 2 update.³ We now prove the following lemma, which will be useful for our primal-dual analysis.

LEMMA 2.1. *For every constraint s , the label $\ell_u^{(s)}$ is monotonically non-decreasing over the course of the algorithm.*

PROOF. Clearly when neither Event 1 or 2 occurs, $\ell_u^{(s)}$ can only increase (this follows directly from the definition of the update procedure). Thus, it suffices to show that $\ell_u^{(s)}$ does not decrease when it participates in Event 1 or 2. In particular, we will show that $\ell_u^{(s)}$ has an identical value after $L(s)$ and $T(s)$ have been modified in either event. We will show that this holds for Event 1, noting that the argument for Event 2 is identical.

Suppose the updates for Event 1 occur for a constraint s and a descendant s' , triggered by a dimension $k \in s$. The changes are: s' is added to $T(s)$, all descendant singleton-budgets of s' that were in $L(s)$ are removed from $L(s)$, and all descendants of s' that were in $T(s)$ are removed from $T(s)$. Recall that before the event, we have that

$$\ell_u^{(s)} = \frac{\sum_{\{k\} \in L(s)} R_u^{(k)}}{B_u^{(s)} - \sum_{w \in T(s)} B_u^{(w)}}$$

$$\ell_u^{(s')} = \frac{\sum_{\{k\} \in L(s')} R_u^{(k)}}{B_u^{(s')} - \sum_{w' \in T(s')} B_u^{(w')}}.$$

Therefore, using the definitions of $L(s)$ and $T(s)$ before they are modified by the event, the new $\ell_u^{(s)}$ (denoted $\ell_{\text{new}}^{(s)}$) can be written as:

$$\ell_{\text{new}}^{(s)} = \frac{\sum_{\{k\} \in L(s)} R_u^{(k)} - \sum_{\{k\} \in L(s')} R_u^{(k)}}{B_u^{(s)} - \sum_{w \in T(s)} B_u^{(w)} - B_u^{(s')} + \sum_{w' \in T(s')} B_u^{(w')}}. \quad (8)$$

Since $\ell_u^{(s)} = \ell_u^{(s')}$ at the moment Event 1 occurs, we have $\ell_{\text{new}}^{(s)} = \ell_u^{(s)} = \ell_u^{(s')}$ (which follows from the fact that $a/b = c/d = \alpha$ implies $(a - c)/(b - d) = \alpha$).

To complete the proof, note that the above argument does not exclude the possibility of $\sum_{\{k\} \in L(s)} R_u^{(k)} = 0$ and the denominator in Eqn. (17) being negative (if this were to happen, the increments to $R_u^{(k)}$ would decrease the label by making it more negative).

³In terms of how $L(s)$ and $T(s)$ are updated, Event 1 is the reverse of Event 2. So, reversing the example in the Figure 3 will provide the reader with an Event 1 example.

However, in both events this cannot be the case. First observe that Event 2 can only occur between two non-zero labels (since in Event 2, $s' \in T(s)$ before the event, which implies a strict inequality $\ell_u^{(s)} < \ell_u^{(s')}$). Event 1 can (and will) occur when $\ell_u^{(s)} = \ell_u^{(s')} = 0$, but in Event 1, the denominator of $\ell_u^{(s')}$ must always be smaller than the denominator of $\ell_u^{(s)}$. This is because Event 1 can only occur between two labels such that $\{k\} \in L(s)$ where k the dimension is currently being incremented. Since $\ell_u^{(s')}$ is surpassing $\ell_u^{(s)}$ at the transition point in Event 1, it must be the case that $\ell_u^{(s')}$ is increasing at a higher rate than $\ell_u^{(s)}$. This implies $\ell_u^{(s')}$ must have a smaller denominator than $\ell_u^{(s)}$ before the modifications to $L(s)$ and $T(s)$. \square

2.3. Algorithm Definition and Analysis

Using our dual formulation and labeling scheme, we are now ready to define and analyze our algorithm. Consider the arrival of impression v . Define $g_u(s) = \max_{s' \in A_s} \ell_u^{(s')}$, i.e., the maximum label of an ancestor of s in the forest F_u (including s itself). Our algorithm assigns impression v to bidder $u = \arg \max_{u' \in U} \{D_{u'v}\}$, where $D_{uv} = \sum_{k \in t_u} (1 - e^{g_u^{(k)}} - 1) r_{uv}^{(k)}$ and t_u is the current active dimensions for bidder u .

For the rest of the section, let $\rho = e/(e-1)$. For a primal assignment of impression v to bidder u , we change the dual solution by setting $\sigma_v = \rho \cdot D_{uv}$ and update $\gamma_u^{(s)}$ to be

$$\gamma_u^{(s)} = \frac{e^{g_u^{(s)}} - 1}{e - 1} = \rho(e^{g_u^{(s)}} - 1),$$

where $g_u^{(s)}$ is computed *after* the assignment of the current impression v .

For the competitive analysis, it suffices to show that a) the ratio between dual and primal objectives is at most ρ , and b) the dual solution is feasible.

Primal-Dual Ratio. Our goal is to show that when an impression v is assigned to a bidder u , the change in dual objective is at most $\rho = e/(e-1)$ times that of the primal objective. First, note that the dual objective is a function of the labels, and we have argued above that the labels do not change when either of Event 1 or 2 happens. Therefore, we only need to account for the change in the dual objective when the labels change but neither of the two events happen. Let us define S_u^* as the subset of constraints in S_u where the value of $\gamma_u^{(s)}$ is different from $\gamma_u^{(p(s))}$:

$$S_u^* = \{s \in S_u : \gamma_u^{(s)} \neq \gamma_u^{(p(s))}\}.$$

We can rewrite the dual objective as $\sum_{v \in V} \sigma_v + \sum_u \sum_{s \in S_u^*} (\gamma_u^{(s)} - \gamma_u^{(p(s))}) B_u^{(s)}$ since for all the other terms, the value of $\gamma_u^{(s)} - \gamma_u^{(p(s))} = 0$. For any constraint $s \in S_u^*$, let $p_u^*(s)$ be its closest ancestor in F_u that is also in S_u^* . Now, observe that by Property 2 of labels and the definition of $g_u^{(s)}$, $T(s) = \{s' \in S_u^* : p_u^*(s') = s\}$ for any $s \in S_u^*$. Then, the dual objective can be further rewritten as

$$\sum_{v \in V} \sigma_v + \sum_u \sum_{s \in S_u^*} \gamma_u^{(s)} \left(B_u^{(s)} - \sum_{s' \in T(s)} B_u^{(s')} \right).$$

As earlier, we will analyze the change in the dual and primal objectives when the revenue on a dimension k is incremented by an infinitesimal amount $\Delta r_{uv}^{(k)}$. Note that for any singleton constraint $\{k\}$, there is a unique $s \in S_u^*$ satisfying $k \in L(s)$; furthermore, $g_u^{(\{k\})} = g_u^{(s)} = \ell_u^{(s)}$. Therefore, the only dual variable in S^* (i.e., in the dual

objective given above) that changes is $\gamma_u^{(s)}$. Let us denote the change in $g_u^{(s)}$ by $\Delta g_u^{(s)}$. Using the small bids assumption, we can write:

$$\begin{aligned} \Delta \gamma_u^{(s)} \cdot \left(B_u^{(s)} - \sum_{s' \in T(s)} B_u^{(s')} \right) &= \frac{\partial \gamma_u^{(s)}}{\partial g_u^{(s)}} \cdot \Delta g_u^{(s)} \cdot \left(B_u^{(s)} - \sum_{s' \in T(s)} B_u^{(s')} \right) \\ &= \rho \cdot (e^{g_u^{(s)}} - 1) \cdot \frac{\Delta r_{uv}^{(k)}}{B_u^{(s)} - \sum_{s' \in T(s)} B_u^{(s')}} \cdot \left(B_u^{(s)} - \sum_{s' \in T(s)} B_u^{(s')} \right) \\ &= \rho \cdot (e^{g_u^{(\{k\})}} - 1) \cdot \Delta r_{uv}^{(k)}. \end{aligned}$$

Summing over all the infinitesimal changes in revenue, the total change in the dual objective for the assignment of impression v is given by $\sigma_v + \rho \cdot \sum_{k \in t_u} (e^{g_u^{(\{k\})}} - 1) \cdot r_{uv}^{(k)}$, where t_u is the set of active dimensions. Since the algorithm sets

$$\sigma_v = \rho \cdot \sum_{k \in t_u} (1 - e^{g_u^{(\{k\})}} - 1) r_{uv}^{(k)},$$

the total change in the dual objective can be written as:

$$\rho \cdot \sum_{k \in t_u} (1 - e^{g_u^{(\{k\})}} - 1) r_{uv}^{(k)} + \rho \cdot \sum_{k \in t_u} (e^{g_u^{(\{k\})}} - 1) \cdot r_{uv}^{(k)} = \rho \cdot \sum_{k \in t_u} r_{uv}^{(k)},$$

which is exactly ρ times the increase in the primal objective.

Dual Feasibility. Finally, we argue that the dual is feasible when the algorithm terminates.

LEMMA 2.2. *At the end of the algorithm, the dual is feasible.*

PROOF. The feasibility of Eqn. (15) follows directly from definition of $g_u^{(s)}$, and the fact $\gamma_u^{(s)}$ is a non-decreasing function of $g_u^{(s)}$. We now show Eqn. (14). Let t_u be the set of active dimensions for bidder u when impression v arrived. First, observe that for all $k \notin t_u$, we have that $g_u^{(\{k\})} = 1$. This follows from the fact that if k is inactive, a constraint $s \in S_u$ containing k has reached $\kappa_u^{(s)} = 1$ (and thus $\ell_u^{(s)} = 1$ as well). Since s is an ancestor of $\{k\}$ in F_u , we also have $g_u^{(\{k\})} = \gamma_u^{(\{k\})} = 1$.

Let u' be the bidder to which the algorithm assigned impression v . Let $\widehat{g}_u^{(s)}$ be the value of $g_u^{(s)}$ when v was assigned (and define $\widehat{g}_{u'}^{(s)}$ similarly). We have the following:

$$\begin{aligned} \sum_k \gamma_u^{(\{k\})} r_{uv}^{(k)} + \sigma_v &= \sum_k \gamma_u^{(\{k\})} r_{uv}^{(k)} + \rho \sum_{k \in t_u} (1 - e^{\widehat{g}_{u'}^{(\{k\})}} - 1) r_{u'v}^{(k)} \quad (\text{by substituting } \sigma_v) \\ &= \sum_{k \notin t_u} r_{uv}^{(k)} + \rho \sum_{k \in t_u} (e^{g_u^{(\{k\})}} - 1/e) r_{uv}^{(k)} + \rho \sum_{k \in t_u} (1 - e^{\widehat{g}_{u'}^{(\{k\})}} - 1) r_{u'v}^{(k)}, \quad (9) \end{aligned}$$

where the second equality follows by substituting $\gamma_u^{(\{k\})}$ and the fact that $\gamma_u^{(\{k\})} = 1$ for all $k \notin t_u$. We can now establish Eqn. (14) as follows:

$$\begin{aligned}
\sum_k \gamma_u^{\{\{k\}\}} r_{uv}^{(k)} + \sigma_v &\geq \sum_{k \notin t_u} r_{uv}^{(k)} + \rho \sum_{k \in t_u} (e^{g_u^{\{\{k\}\}} - 1} - 1/e) r_{uv}^{(k)} + \rho \sum_{k \in t_u} (1 - e^{\widehat{g_u^{\{\{k\}\}} - 1}}) r_{u'v}^{(k)} \\
&\geq \sum_{k \notin t_u} r_{uv}^{(k)} + \rho \sum_{k \in t_u} (e^{g_u^{\{\{k\}\}} - 1} - 1/e) r_{uv}^{(k)} + \rho \sum_{k \in t_u} (1 - e^{g_u^{\{\{k\}\}} - 1}) r_{uv}^{(k)} \\
&\geq \sum_{k \notin t_u} r_{uv}^{(k)} + \rho \sum_{k \in t_u} (1 - 1/e) r_{uv}^{(k)} = \sum_k r_{uv}^{(k)} \quad (\text{since } \rho = e/(e-1)).
\end{aligned}$$

The first inequality follows from Eqn. (18) and the fact that the algorithm assigns v to $u' = \arg \max_{u \in U} \{D_{uv}\}$. The second inequality is because $1 - e^{g_u^{\{\{k\}\}} - 1}$ is a non-increasing function of $g_u^{\{\{k\}\}}$, and the third inequality follows since $r_{uv}^{(k)} \geq r_{u'v}^{(k)}$. \square

3. ADWORDS WITH GENERAL BUDGET CONSTRAINTS (ADGENERAL)

Recall the ADGENERAL problem: we given a set of offline bidders U and a set of impressions V that arrive online, where each bidder-impression pair (u, v) is specified by a bid value $r_{uv}^{(k)}$ for each dimension $k \in K$. The revenue generated from a bidder u is subject to an arbitrary set of budget constraints S_u , where each constraint $s \in S_u$ caps the total revenue generated from a subset of dimensions K_s to a given budget $B_u^{(s)}$.

In this section, we will prove an $O(\lg p)$ upper bound for ADGENERAL (Theorem 1.1), where $p = \max_{u \in U, k \in K} |\{s \in S_u : k \in K_s\}|$ denotes the the maximum number of budget constraints to which any dimension belongs.

3.1. Algorithm Definition

As in Section ??, let $\kappa_u^{(s)}$ denote the current utilization of budget $B_u^{(s)}$. The algorithm (we call it ALGO) uses an exponential potential function defined by:

$$\phi = \sum_u \sum_s \phi_u^{(s)} = \sum_u \sum_s \frac{B_u^{(s)}}{p} \left((2p+2)^{\kappa_u^{(s)}} - 1 \right),$$

where $\kappa_u^{(s)}$ is defined as the fraction of $B_u^{(s)}$ that has already been used by the algorithm at any stage. Note that $\phi = 0$ initially.

At any stage of ALGO, a dimension k is said to be *active* for bidder u if and only if $\sum_{s: k \in s} \frac{\phi_u^{(s)}}{B_u^{(s)}} \leq 1$; otherwise, dimension k is said to be *inactive* for bidder u . (Note that this is a different definition of active dimensions than what is used in Section ??). ALGO only attempts to earn revenue on active dimensions, and hence, the total revenue if impression v is allocated to bidder u is given by:

$$r_{uv} = \sum_{k \in A_u} r_{uv}^{(k)}, \text{ where } A_u \text{ is the set of current active dimensions for bidder } u.$$

The algorithm makes a greedy assignment with respect to r_{uv} , i.e., it assigns impression v to $\arg \max_u r_{uv}$. Note that it is possible that $A_u = \emptyset$ for all bidders u , and therefore the algorithm does not assign impression v to any bidder, even though there are dimensions and bidders where it could have earned revenue. This completes the description of our algorithm.

3.2. Algorithm Analysis

For our analysis, it will be sufficient to quantify the small bids assumption as the following property for any impression v , bidder u , dimension k , and constraint s such

that $k \in s$:

$$\sum_{k \in s} \frac{r_{uv}^{(k)}}{B_u^{(s)}} \leq \frac{1}{\lg(2p+2)}. \quad (10)$$

The following lemma establishes the feasibility of the algorithm's solution. We relegate its proof to the full version of the paper due to space limitations, but it follows almost directly from the definition of active dimensions.

LEMMA 3.1. *If ALGO assigns an impression v to a bidder u , then it can earn revenue on all the active dimensions A_u of u without violating any constraint.*

This lemma implies that ALGO is indeed able to earn revenue on all active dimensions of a bidder u when it assigns an impression to u .

Next, we will bound the total revenue of an optimal solution that we denote by OPT against the total revenue of ALGO. Let $u_{\text{OPT}}(v)$ (resp., $u_{\text{ALGO}}(v)$) be the bidder that OPT (resp., ALGO) allocates impression v to. For every dimension k that OPT earns revenue from, one of the following holds:

- (1) **Case 1:** dimension k is active for bidder $u_{\text{OPT}}(v)$ in ALGO when impression v arrives, and ALGO assigns v to the same bidder, i.e., $u_{\text{ALGO}}(v) = u_{\text{OPT}}(v)$.
- (2) **Case 2:** dimension k is active for bidder $u_{\text{OPT}}(v)$ in ALGO when impression v arrives, but ALGO assigns v to a different bidder, i.e., $u_{\text{ALGO}}(v) \neq u_{\text{OPT}}(v)$.
- (3) **Case 3:** dimension k is inactive for bidder $u_{\text{OPT}}(v)$ in ALGO when impression v arrives.

We partition the dimensions that OPT earns revenue from into active and inactive dimensions (according to their status in ALGO for bidder $u_{\text{OPT}}(v)$). For active dimensions, the next lemma gives a straightforward charging argument using the greediness of the choice made by ALGO.

LEMMA 3.2. *For any impression v , the total revenue earned by OPT on the active dimensions is at most the total revenue earned by ALGO overall.*

PROOF. For case 1 above (OPT and ALGO choose the same bidder), the two algorithms earn the same revenue on the active dimensions. For case 2 above, the fact ALGO makes a greedy choice implies that it earns at least as much revenue by assigning to different bidder as it would have made by assigning to u , which includes the revenue on all the active dimensions. \square

The more involved case is that of inactive dimensions. In this case, we use a different global charging argument over all dimensions, based on the potential function. In particular, we show that for a bidder u , the total revenue of OPT from inactive dimensions (recall that this only includes revenue from impressions that arrived *after* the dimension became inactive in ALGO) can be charged, up to a logarithmic loss, to the revenue that ALGO earned overall from bidder u .

LEMMA 3.3. *Fix a bidder u . The total revenue that OPT earns in inactive dimensions for bidder u is at most the final potential of bidder u in ALGO.*

PROOF. For any dimension k , let $v \in V_{u,k}$ denote the subset of impressions assigned to u by OPT that arrived after k became an inactive dimension for bidder u in ALGO. We need to bound the total revenue earned by OPT on dimension k from impressions in $V_{u,k}$, summed over all k . For any impression $v \in V_{u,k}$, we have $\sum_{s:k \in s} \frac{\phi_u^{(s)}}{B_u^{(s)}} > 1$, where $\phi_u^{(s)}$ is the final potential for constraint s of bidder u . Thus, the revenue that OPT

earns from impressions $v \in V_{u,k}$ on dimension k , summed over all dimensions, can be bounded as follows:

$$\sum_k \sum_{v \in V_{u,k}} r_{uv}^{(k)} < \sum_k \sum_{v \in V_{u,k}} r_{uv}^{(k)} \sum_{s:k \in s} \frac{\phi_u^{(s)}}{B_u^{(s)}} = \sum_s \phi_u^{(s)} \sum_{k \in s} \sum_{v \in V_{u,k}} \frac{r_{uv}^{(k)}}{B_u^{(s)}} \leq \sum_s \phi_u^{(s)},$$

where the last inequality follows from the feasibility of OPT. \square

Finally, we need to lower bound the final potential of a bidder in terms of the revenue it generates for ALGO. This is done in the following lemma.

LEMMA 3.4. *The increase in potential of a bidder u during the course of ALGO is at most $4 \lg(2p+2)$ times the revenue that ALGO earns from u .*

PROOF. Suppose ALGO assigns impression v to bidder u . Let K_a denote the set of active dimensions for bidder u when this assignment is made. Let $\eta_u^{(s)} = \sum_{k:k \in K_a \cap s} \frac{r_{uv}^{(k)}}{B_u^{(s)}}$. Then, the increase in potential is given by:

$$\begin{aligned} \sum_s \Delta \phi_u^{(s)} &= \sum_s \frac{B_u^{(s)}}{p} \left((2p+2)^{\kappa_u^{(s)} + \eta_u^{(s)}} - (2p+2)^{\kappa_u^{(s)}} \right) = \sum_s \frac{B_u^{(s)}}{p} (2p+2)^{\kappa_u^{(s)}} \left((2p+2)^{\eta_u^{(s)}} - 1 \right) \\ &= \sum_s \left(\phi_u^{(s)} + \frac{B_u^{(s)}}{p} \right) \cdot \left(2^{\eta_u^{(s)} \cdot \lg(2p+2)} - 1 \right) \\ &\leq \sum_s \left(\phi_u^{(s)} + \frac{B_u^{(s)}}{p} \right) \cdot 2 \lg(2p+2) \cdot \sum_{k:k \in K_a \cap s} \frac{r_{uv}^{(k)}}{B_u^{(s)}}, \end{aligned} \quad (11)$$

where the last inequality follows since $\eta_u^{(s)} = \sum_{k:k \in K_a \cap s} \frac{r_{uv}^{(k)}}{B_u^{(s)}}$ and $a^x \leq 1 + ax$ for $0 \leq x \leq 1, a \geq 1$. By rearranging the RHS of inequality (20), we have:

$$\sum_s \Delta \phi_u^{(s)} \leq 2 \lg(2p+2) \cdot \sum_{k \in K_a} r_{uv}^{(k)} \cdot \left(\sum_{s:k \in s} \frac{\phi_u^{(s)}}{B_u^{(s)}} + \frac{|\{s : k \in s\}|}{p} \right) \leq 4 \lg(2p+2) \cdot \sum_{k \in K_a} r_{uv}^{(k)},$$

since k is active and $p \geq |\{s : k \in s\}|$. \square

A competitive ratio of $O(\lg p)$ for ADGENERAL in the small bids case now follows from Lemmas 3.1, 3.3, and 3.4, completing proof of the upper bound in Theorem 1.1.

REFERENCES

- Shalinda Adikari and Kaushik Dutta. 2015. Real Time Bidding in Online Digital Advertisement. In *DESRIST*. 19–38.
- Gagan Aggarwal, Yang Cai, Aranyak Mehta, and George Pierrakos. 2014. Biobjective Online Bipartite Matching. In *WINE*. 218–231.
- Gagan Aggarwal, Gagan Goel, Chinmay Karande, and Aranyak Mehta. 2011. Online Vertex-Weighted Bipartite Matching and Single-bid Budgeted Allocations. In *SODA*. 1253–1264.
- Shipra Agrawal and Nikhil R. Devanur. 2015. Fast Algorithms for Online Stochastic Convex Programming. In *SODA*. 1405–1424.
- Baruch Awerbuch, Yossi Azar, and Serge A. Plotkin. 1993. Throughput-Competitive On-Line Routing. In *STOC*. 32–40.
- Joel Barajas Zamora. 2015. Online Display Advertising Causal Attribution and Evaluation. (2015).

- Vijay Bharadwaj, Peiji Chen, Wenjing Ma, Chandrashekar Nagarajan, John Tomlin, Sergei Vassilvitskii, Erik Vee, and Jian Yang. 2012. SHALE: an efficient algorithm for allocation of guaranteed display advertising. In *SIGKDD*. 1195–1203.
- Benjamin E. Birnbaum and Claire Mathieu. 2008. On-line bipartite matching made simple. *SIGACT News* 39, 1 (2008), 80–87.
- Allan Borodin and Ran El-Yaniv. 1998. *Online computation and competitive analysis*. Cambridge University Press.
- Niv Buchbinder, Kamal Jain, and Joseph Naor. 2007. Online Primal-Dual Algorithms for Maximizing Ad-Auctions Revenue. In *ESA*. 253–264.
- Richard E. Chatwin. 2013. An overview of computational challenges in online advertising. In *American Control Conference, ACC*. 5990–6007.
- Nikhil R. Devanur and Kamal Jain. 2012. Online matching with concave returns. In *STOC*. 137–144.
- Nikhil R. Devanur, Kamal Jain, and Robert D. Kleinberg. 2013. Randomized Primal-Dual analysis of RANKING for Online BiPartite Matching. In *SODA*. 101–107.
- Jon Feldman, Nitish Korula, Vahab S. Mirrokni, S. Muthukrishnan, and Martin Pál. 2009. Online Ad Assignment with Free Disposal. In *WINE*. 374–385.
- Arpita Ghosh, Randolph Preston McAfee, Kishore Papineni, and Sergei Vassilvitskii. 2009. Bidding for Representative Allocations for Display Advertising. In *WINE*. 208–219.
- Gagan Goel and Aranyak Mehta. 2008. Online budgeted matching in random input models with applications to Adwords. In *SODA*. 982–991.
- S. Ali Hojjat, John Turner, Suleyman Cetintas, and Jian Yang. 2014. Delivering Guaranteed Display Ads under Reach and Frequency Requirements. In *AAAI*. 2278–2284.
- Bala Kalyanasundaram and Kirk Pruhs. 2000. An optimal deterministic algorithm for online b-matching. *Theor. Comput. Sci.* 233, 1-2 (2000), 319–325.
- Richard M. Karp, Umesh V. Vazirani, and Vijay V. Vazirani. 1990. An Optimal Algorithm for On-line Bipartite Matching. In *STOC*. 352–358.
- Aranyak Mehta. 2013. Online Matching and Ad Allocation. *Foundations and Trends in Theoretical Computer Science* 8, 4 (2013), 265–368.
- Aranyak Mehta and Debmalya Panigrahi. 2012. Online Matching with Stochastic Rewards. In *FOCS*. 728–737.
- Aranyak Mehta, Amin Saberi, Umesh V. Vazirani, and Vijay V. Vazirani. 2007. AdWords and generalized online matching. *J. ACM* 54, 5 (2007).
- Aranyak Mehta, Bo Waggoner, and Morteza Zadimoghaddam. 2015. Online Stochastic Matching with Unequal Probabilities. In *SODA*. 1388–1404.