A Truthful Learning Mechanism for Contextual Multi–Slot Sponsored Search Auctions with Externalities

Nicola Gatti, Alessandro Lazaric, Francesco Trov'o

A Truthful Learning Mechanism for Contextual Multi–Slot Sponsored Search Auctions with Externalities

NICOLA GATTI, Politecnico di Milano
ALESSANDRO LAZARIC, INRIA Lille – Nord Euopre
FRANCESCO TROVÒ, Politecnico di Milano

Sponsored search auctions constitute one of the most successful applications of *microeconomic mechanisms*. In mechanism design, auctions are usually designed to incentivize advertisers to bid their truthful valuations and, at the same time, to assure both the advertisers and the auctioneer a non–negative utility. Nonetheless, in sponsored search auctions, the click–through–rates (CTRs) of the advertisers are often unknown to the auctioneer and thus standard incentive compatible mechanisms cannot be directly applied and must be paired with an effective learning algorithm for the estimation of the CTRs. This introduces the critical problem of designing a learning mechanism able to estimate the CTRs as the same time as implementing a truthful mechanism with a revenue loss as small as possible compared to an optimal mechanism designed with the true CTRs. Previous works showed that in single–slot auctions the problem can be solved using a suitable exploration–exploitation mechanism able to achieve a per–step regret of order $O(T^{-1/3})$ (where T is the number of times the auction is repeated). In this paper we extend these results to the general case of contextual multi–slot auctions with position– and ad–dependent externalities. In particular, we prove novel upper–bounds on the revenue loss w.r.t. to a VCG auction and we report numerical simulations investigating their accuracy in predicting the dependency of the regret on the number of rounds T, the number of slots K, and the number of advertisements n.

Categories and Subject Descriptors: I.2 [**Artificial Intelligence**]: Miscellaneous

General Terms: Algorithms, Economics, Theory

Additional Key Words and Phrases: Sponsored search auction, Multi–arm bandit, Learning mechanism

1. INTRODUCTION

Sponsored search auctions (SSAs) constitute one of the most successful applications of *microeconomic mechanisms*, producing a revenue of about $6 billion dollars in the US alone in the first half of 2010 [IAB 2010]. In a SSA, a number of *advertisers* (from here on *advs*) bid to have their *sponsored links* (from here on *ads*) displayed in some slot alongside the search results of a keyword. Sponsored search auctions adopt a *pay–per–click* scheme, requiring positive payments to an adv only if its ad has been clicked. Given an allocation of ads over the slots, each ad is associated with a *click–through–rate* (CTR) defined as the probability that such ad will be clicked by the user. CTRs are estimated by the auctioneer and play a crucial role in the auction, since they are used by the auctioneer to find the optimal allocation (in expectation) and to compute the payments for each ad.

There is a large number of works formalizing SSAs as a *mechanism design* problem [Narahari et al. 2009], where the objective is to design an auction mechanism

---

A short version of this paper (2–page extended abstract) has been accepted for publication at the 11th International Conference on Autonomous Agents and Multi–Agent Systems (AAMAS'12).

EC'12, June 4–8, 2012, Valencia, Spain.

that incentivizes advs to bid their truthful valuations (needed for stability) and that assures both the advs and the auctioneer to have a non–negative utility. The most common SSA mechanism is the *generalized second price* (GSP) auction [Edelman et al. 2007; Varian 2007]. This mechanism is proved not to be incentive compatible (advs may gain more by bidding non–truthful valuations) and different bidding strategies are investigated in [Edelman et al. 2007]. A generalization of the Vickrey–Clarke–Groves (VCG) mechanism (assuring incentive compatibility) for SSAs has been investigated in [Narahari et al. 2009]. Although the VCG mechanism is not currently adopted by the search engines, a number of scientific theoretical results builds upon it. Interestingly, there is a strict relationship between GSP and VCG: the worst (for the auctioneer) *ex post* equilibrium with GSP is payoff–equivalent to the truthful equilibrium with the VCG [Edelman et al. 2007]. This implies that any result derived for the VCG mechanism also applies to a GSP whenever the bidders converge to an equilibrium. Other works focus instead on modeling the user behavior. Among these models, the one that better predicts the human behavior and is most commonly used, is the *cascade* model, which assumes that the user scans the links from the top to the bottom in a Markovian way [Aggarwal et al. 2008; Kempe and Mahdian 2008]. These models introduce negative externalities in the auction whereby the click probability, and therefore the adv's profit, depends on which ads are displayed at the same time on the other slots.

In this paper, we focus on the problem of how to design a truthful mechanism when the CTRs are not known and need to be estimated. This problem is particularly relevant in practice because the assumption that all the CTRs are known beforehand is rarely realistic. Furthermore, it also poses interesting scientific challenges since it represents one of the first examples where learning theory is paired with mechanism design techniques to obtain effective methods to learn under equilibrium constraints (notably the incentive compatibility property). The problem of estimating the CTRs and to identify the best allocation of ads is effectively formalized as a *multi–arm bandit problem* [Robbins 1952] where each ad is an arm and the objective is to minimize the cumulative regret (i.e., the revenue loss w.r.t. an optimal allocation defined according to the true CTRs). The problem of budgeted advs (i.e., auctions where the total amount of money each adv is willing to pay is limited) with multiple queries is considered in [Pandey and Olston 2006]. This problem is formalized as a budgeted multi–bandit multi–arm problem, where each bandit corresponds to a query, and an algorithm is proposed with explicit bounds over the regret on the revenue. Nonetheless, the proposed method works in a non–strategic environment, where the advs do not try to influence the outcome of the auction and always bid their true values. The strategic dimension of SSAs is partially taken into consideration in [Langford et al. 2010] where the advs are assumed to play a bidding strategy at the equilibrium w.r.t. a set of estimated CTRs which are available to both the auctioneer and the advs. The authors introduce a learning algorithm which explores different rankings on the ads so as to improve the CTR estimates and, at the same, not to introduce incentives for the advs to deviate from the previous equilibrium strategy. A more complete notion of truthfulness for bandit algorithms in multi–slot SSAs is studied in [Gonen and Pavlov 2007b]. In particular, they build on the action elimination algorithm in [Even-Dar et al. 2006] and they report a probably approximately correct (PAC) analysis of its performance. Unfortunately, as pointed in [Devanur and Kakade 2009] and [Babaioff et al. 2008] the mechanism is not guaranteed to be truthful and thus it only works when the advertisers bid their true values. An extension to the action elimination algorithm is also proposed in [Gonen and Pavlov 2007a] for the more general setting where budgeted advs are allowed to enter and exit the auction at different time instants that they declare along with their bid. The authors derive an algorithm that approximately achieves the best social welfare under the assumption that the gain of untruthful declarations is limited. Fi-

nally, single–slot online advertising is studied also in [Nazerzadeh et al. 2008] where the notion of Bayesian incentive compatibility (BIC) is taken into consideration and an asymptotically BIC and *ex ante* efficient mechanism is introduced. The most complete study of truthful bandit mechanisms so far is reported in [Devanur and Kakade 2009] and [Babaioff et al. 2008]. These recent works first provided a complete analysis on the constraints truthfulness forces on the multi–arm bandit algorithm, showing that no truthful bandit mechanism can achieve a regret smaller than $\tilde{\Omega}(T^{2/3})$. Furthermore, they also suggest nearly–optimal algorithms for the simple case of single–slot SSAs.

In this paper, we build on the exploration–exploitation mechanisms introduced and analyzed in [Devanur and Kakade 2009] and [Babaioff et al. 2008] and we report a regret analysis for an adaptive VCG mechanism for the general case of contextual multi–slot auctions with externalities. More precisely, the main contributions of the present paper can be summarized as follows:

Multi–slot auctions with position–dependent externalities. We extend the existing exploration–exploitation mechanism to multi–slot auctions and we derive a regret bound for position–dependent externalities. The main finding is that previous single–slot results smoothly extend to the multi–slot case and the regret R_T has a sub–linear dependency on the number of slots K, i.e., $R_T = \tilde{O}(T^{2/3}n^{1/3}K^{2/3})$. Numerical simulations confirm the accuracy of the bound (in a worst–case scenario w.r.t. the qualities).

Multi–slot auctions with position/ad–dependent externalities. We derive regret bounds for the general case of position/ad–dependent externalities in which CTRs depend on the ads' allocation. In this case, the regret analysis is more complicated and needs different steps than in the position–dependent case. The regret bound shows that the learning problem is more difficult and the regret now scales as $\tilde{O}(T^{2/3}K^{2/3}n)$. Nonetheless, numerical simulations suggest that this dependency might be over–estimated and we conjecture that a more accurate bound would be $\tilde{O}(T^{2/3}K^{4/3}n^{1/3})$, thus implying a worse dependency in the number of slots w.r.t. to the position–dependent case.[1]

Contextual multi–slot auctions. Finally, we report a regret analysis for auctions which are characterized by a context x summarizing information such as user's profile, webpage content, etc. The resulting bound displays similar characteristics as for the no–context position/ad–dependent externalities case.

The paper is organized as follows. In Section 2 we introduce the notation and the learning mechanism problem. From Section 3 to Section 5 we report and discuss the main regret bounds and in Section 6 we analyze simple numerical simulations to test the accuracy of the theoretical bounds. Section 7 concludes the paper and proposes future directions of investigation. The detailed proofs of the theorems are reported in [A. Lazaric 2012].

2. NOTATION AND BACKGROUND

In this section we introduce the basic notation used throughout the rest of the paper, we review the definition of the VGC mechanism for SSAs, and we report the main technical results available for the learning problem.

We consider a standard model of SSAs. We denote by $\mathcal{N} = \{1, \dots, n\}$ the set of ads ($i \in \mathcal{N}$ is a generic ad) and by a_i the adv of ad i (we assume each adv has only one ad). Each ad i is characterized by a *quality* ρ_i, defined as the probability that i is clicked once observed by the user, and by a *value* $v_i \in \mathcal{V}$, with $\mathcal{V} = [0, V]$, that a_i receives once i is clicked (a_i receives a value of zero if not clicked). While qualities ρ_i are common

[1]Notice that $\tilde{O}(T^{2/3}K^{4/3}n^{1/3}) \leq \tilde{O}(T^{2/3}K^{2/3}n)$, since $K \leq n$.

knowledge, values v_i are private information of the advs. We denote by $\mathcal{K} = \{1, \ldots, K\}$, with $K < n$,[2] the set of available slot (k denotes a generic slot). Although at each round only K ads can be actually displayed, for notational convenience we define an ad–slot allocation rule α as a full bijective mapping from n ads to n slots (i.e., $\alpha : \mathcal{N} \to \mathcal{N}$) such that $\alpha(i) = k$ if ad $i \in \mathcal{N}$ is displayed at slot k. We assume that for all the non–allocated ads, $\alpha(i)$ takes an arbitrary value from $K+1$ to n so as to preserve the bijectivity of α. We also define the inverse slot–ad allocation rule $\beta = \alpha^{-1}$ such that $\beta(k) = i$ if slot k displays ad i (i.e., $\alpha(i) = k$). We denote by $\mathcal{A}$ and $\mathcal{B}$ the set of all the possible ad–slot and slot–ad mappings respectively. Finally, we define $\mathcal{A}_{-i} = \{\alpha \in \mathcal{A}, \alpha(i) = n\}$ as the set of allocations where ad i is never displayed.

In order to describe the user's behavior, we adopt the popular cascade model defined by [Kempe and Mahdian 2008; Aggarwal et al. 2008]. The user is assumed to look through the list of slots from the top to the bottom and the probability that the user observes the next ad i depends on the slot in which i is displayed at (*position–dependent externalities*) and/or on the ad that precedes i in the allocation (*ad–dependent externalities*). We define the discount factor $\gamma_k(i)$ as the probability that a user observing ad i in the slot $k-1$ will observe the ad in the next slot (γ_1 is set to 1 by definition). The cumulative discount factors $\Gamma_k(\beta)$, i.e., the probability that a user observes the ad displayed at slot k given a slot–ad allocation β, is defined as:

$$\Gamma_k(\beta) = \begin{cases} 1 & \text{if } k = 1 \\ \prod_{l=2}^{k} \gamma_l(\beta(l-1)) & \text{if } 2 \leq k \leq K \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

With abuse of notation, we use interchangeably $\Gamma_k(\beta)$ and $\Gamma_k(\alpha)$ (for $\beta = \alpha^{-1}$). Given an allocation rule α, $\Gamma_{\alpha(i)}(\alpha)\rho_i$ is the *click through rate* (CTR), representing the probability of ad i to be clicked. We notice that, according to this model, the user might click multiple ads at each impression. As a result, the *expected value* (w.r.t. the user's click) of adv a_i for an allocation α is $\Gamma_{\alpha(i)}(\alpha)\rho_i v_i$. Finally, we define the *social welfare* of an allocation α (equivalently, $\beta = \alpha^{-1}$) as the cumulative advs' expected values

$$\text{SW}(\alpha) = \text{SW}(\beta) = \sum_{i=1}^{n} \Gamma_{\alpha(i)}(\alpha)\rho_i v_i = \sum_{k=1}^{n} \Gamma_k(\beta)\rho_{\beta(k)} v_{\beta(k)}.$$

At each round, advs submit bids $\{\hat{v}_i\}_{i \in \mathcal{N}}$ and the auction defines an allocation rule α and payment functions $p_i(\hat{v}_1, \ldots, \hat{v}_n)$. The expected utility of adv a_i is defined as $\Gamma_{\alpha(i)}(\alpha)\rho_i v_i - p_i(\hat{v}_1, \ldots, \hat{v}_n)$. Each adv is an expected utility maximizer and therefore, if it can gain more by misreporting its value (i.e., $\hat{v}_i \neq v_i$), it will do that. Mechanism design aims at finding α and $\{p_i\}_{i \in \mathcal{N}}$ such that some properties are satisfied (see [Mas-Colell et al. 1995] for more formal and detailed definitions):

Incentive Compatibility (IC). This assures that no adv can increase its utility by misreporting its true value. This property is necessary to have stable mechanisms.

Individual Rationality (IR). This assures that each adv receives a non–negative utility in taking part to the auction (given truthful reporting).

Weak Budget Balance (WBB). This assures that the auctioneer has always a non–negative revenue.

Allocative Efficiency (AE). This assures that the allocation maximizes the social welfare.

The unique mechanism for SSAs satisfying the above properties is the Vickrey–Clark–Groves mechanism (VCG) where ads are allocated according to the *efficient* al-

[2]Although $K < n$ is the most common case, the results could be smoothly extended to $K > n$.

location α^*, i.e., the the social welfare maximizer:

$$\alpha^* = \arg\max_{\alpha \in \mathcal{A}} \mathrm{SW}(\alpha). \tag{2}$$

and payments are defined as

$$p_i = \mathrm{SW}(\alpha^*_{-i}) - \mathrm{SW}_{-i}(\alpha^*), \tag{3}$$

where $\mathrm{SW}_{-i}(\alpha) = \mathrm{SW}(\alpha) - \Gamma_{\alpha(i)}(\alpha)\rho_i v_i$ and α^*_{-i} is the efficient allocation in $\mathcal{A}_{-i}$. In words, the payment for adv a_i is the difference between the social welfare that could be obtained by an efficient allocation α^*_{-i} computed removing ad i and the social welfare of the efficient allocation α^* without the contribution of adv a_i. This mechanism is IR in expectation, but not for every possible realization (an adv may have a positive payment even when its ad has not been clicked). Anyway, the mechanism can be easily modified to satisfy IR for every possible realization by using click–contingent payments that are zero if the ad is not clicked and

$$\tilde{p}_i = \frac{\mathrm{SW}(\alpha^*_{-i}) - \mathrm{SW}_{-i}(\alpha^*)}{\Gamma_{\alpha(i)}(\alpha)\rho_i}, \tag{4}$$

if the ad is clicked (so that $\mathbb{E}[\tilde{p}_i] = p_i$).

In many practical problems, the qualities ρ_i are not known in advance and must be estimated at the same time as the auction is deployed. This introduces a tradeoff between *exploring* different possible allocations so as to collect information about the quality of the advs and *exploiting* the estimated qualities so as to implement a truthful high–revenue auction (i.e., a VCG mechanism). This problem could be easily casted as a multi–arm bandit problem [Robbins 1952] and standard techniques could be used to solve it (see e.g., [Auer et al. 2002]). Nonetheless, such an approach would completely overlook the strategic dimension of the problem where advs may choose their bids so as to influence the outcome of the auction and increase their utility. As a result, here we face the more challenging problem where the exploration–exploitation dilemma must be solved so as to maximize the revenue of the auction under the hard constraint of incentive compatibility. Let $\mathfrak{A}$ be an IC mechanism run over T rounds. At each round t, $\mathfrak{A}$ defines an allocation $\hat{\alpha}_t$ and prescribes an expected payment p_{it} for each ad i. The objective of $\mathfrak{A}$ is to obtain a revenue as close as possible to a VCG mechanism computed on the true qualities $\{\rho_i\}_{i \in \mathcal{N}}$.[3] More precisely, we measure the performance of $\mathfrak{A}$ as its cumulative regret over T rounds:

$$\mathcal{R}_T(\mathfrak{A}) = T \sum_{i=1}^{n} p_i - \sum_{t=1}^{T} \sum_{i=1}^{n} p_{it},$$

where p_i is as defined in (3). We notice that the regret does not compare the *actual* payments asked on a specific sequence of clicks ($\tilde{p}_{it}$) but the expected payments p_{it}. Furthermore, since the learning mechanism $\mathfrak{A}$ estimates the qualities from the observed (random) clicks, the expected payments p_{it} are random as well. Thus, in the following we will study the expected regret

$$R_T(\mathfrak{A}) = \mathbb{E}[\mathcal{R}_T(\mathfrak{A})], \tag{5}$$

where the expectation is taken w.r.t. random sequences of clicks. The mechanism $\mathfrak{A}$ is a *no–regret* mechanism if its per–round regret decreases to 0 as T increases, i.e., $\lim_{T \to \infty} R_T/T = 0$. Another popular definition of performance [Gonen and Pavlov

[3] We refer to reader to Appendix D in [A. Lazaric 2012] for a slightly different definition of regret measuring the deviation from the revenue of a VGC mechanism.

Input: Length of exploration phase τ, confidence δ, cumulative factors $\{\Gamma_k(\cdot)\}_{k\in\mathcal{K}}$

Exploration phase
for $t = 1, \dots, \tau$ **do**
  Allocate ads according to an arbitrary allocation rule α_t
  Set the payments $\hat{p}_i = 0$
  Observe the clicks $\{c^i_{\alpha_t(i)}(t)\}_{i=1}^n$
end for
Compute the estimated quality $\hat{\rho}_i = \frac{1}{S_i} \sum_{s=1}^{S_i} \frac{c^i_{\alpha_t(i)}(t)}{\Gamma_{\alpha_t(i)}}$
Compute $\hat{\rho}_i^+ = \hat{\rho}_i + \eta$ (see (8) and (12))

Exploitation phase
for $t = \tau + 1, \dots, T$ **do**
  Allocate ads according to $\hat{\alpha}$ (see Sect. 3 and 4)
  if Ad i is clicked **then**
    Ask for payment $\tilde{p}_i$ (see (9) and (13))
  end if
end for

Fig. 1. Pseudo–code for the A–VCG mechanism.

2007b; Babaioff et al. 2008] is the social welfare regret, measured as the difference between the social welfare of the optimal allocation α^* and of the best estimated allocation $\hat{\alpha}$ (i.e., $\mathrm{SW}(\alpha^*) - \mathrm{SW}(\hat{\alpha})$). We notice that minimizing the social welfare regret does not coincide with minimizing R_T. In fact, once the quality estimates are accurate enough, $\hat{\alpha}_t$ is equal to α^*, and the social welfare regret drops to zero. On the other hand, since p_{it} is defined according to the estimated qualities, even if $\hat{\alpha}_t = \alpha^*$, R_T might still be positive.

The properties required to have a truthful mechanism in single–slot auctions ($K = 1$) are studied in [Devanur and Kakade 2009] and it is shown that any learning algorithm must split the exploration and the exploration in two separate phases in order to incentivize advs to bid their true values. This condition has a strong impact on the regret of the mechanism. In fact, while in a standard bandit problem the distribution–free regret is of order $\Omega(T^{1/2})$, in single–slot auctions, truthful mechanisms cannot achieve a regret smaller than $\Omega(T^{2/3})$. In [Devanur and Kakade 2009] a truthful learning mechanism is designed with a nearly optimal regret of order $\tilde{O}(T^{2/3})$.[4] Similar structural properties for truthful mechanisms are also studied in [Babaioff et al. 2008] and lower–bounds are derived for the social welfare regret. In this paper we build on these theoretical results for truthful mechanisms and we extend the exploration–exploitation algorithm in [Devanur and Kakade 2009] to the general case of contextual multi–slot auctions with position– and position/ad–dependent externalities.

3. POSITION–DEPENDENT EXTERNALITIES

In this section we consider multi–slot auctions with position–dependent externalities where the discount factors γ_k, and the corresponding cumulative factors Γ_k (see equation (1)), do not depend on the ads displayed in the previous slots but only on the position k (i.e., $\gamma_k(\beta) = \gamma_k$). In this case, the efficient allocation α^* (β^*) can be easily computed by allocating the ads to the slots in decreasing order w.r.t. their expected

[4]The $\tilde{O}$ notation hides both constant and logarithmic factors, that is $R_T \leq \tilde{O}(T^{2/3})$ if there exist a and b such that $R_T \leq aT^{2/3} \log^b T$.

value $\rho_i v_i$. More precisely, for any k, let $\max_i(\rho_i v_i; k)$ be the operator returning the k-th largest value in the set, then β^* is such that $\beta^*(k) = \arg\max_i(\rho_i v_i; k)$. This condition also simplifies the definition of the efficient allocation α^*_{-i} when i is removed from $\mathcal{N}$. In fact, for any $j \in \mathcal{N}$, if $\alpha^*(j) < \alpha^*(i)$ (i.e., ad j is displayed before i) then $\alpha^*_{-i}(j) = \alpha^*(j)$, while if $\alpha^*(j) > \alpha^*(i)$ then $\alpha^*_{-i}(j) = \alpha^*(j) - 1$ (i.e., ad j is moved one slot upward), and $\alpha^*_{-i}(i) = n$. By recalling the definition of payment in (3), in case of position–dependent externalities the payment for the adv in slot $k \leq K$ reduces to

$$p_{\beta^*(k)} = \sum_{l=k+1}^{K+1} (\Gamma_{l-1} - \Gamma_l) \max_i(\rho_i v_i; l), \tag{6}$$

while it is equal to 0 for any $k > K$.

Similar to [Devanur and Kakade 2009], we define an exploration–exploitation algorithm to approximate the VCG. The algorithm receives as input the cumulative factors Γ_k and it estimates the quality of each adv during a pure exploration phase of length τ when all the payments are set to 0. Then, quality estimates are used to set up a VCG for all the remaining $T - \tau$ rounds. Unlike the single–slot case, here we can exploit the fact that each ad i has a non-zero CTR $\Gamma_{\alpha(i)}\rho_i$ whenever it is allocated to a slot $\alpha(i) \leq K$. As a result, at each round, we can collect K samples (click or not–click events), one from each slot. Let α_t (for $t \leq \tau$) be an explorative allocation rule defined in an arbitrary way completely independent from the bids. The number of samples collected for each ad i is $S_i = \sum_{t=1}^{\tau} \mathbb{I}\{\alpha_t(i) \leq K\}$. We denote by $c^i_{\alpha_t(i)}(t) \in \{0, 1\}$ the click–event at time t for ad i when displayed at slot $\alpha_t(i)$. Depending on the slot we have different CTRs, thus we reweigh each sample by the cumulative discount factor of the slot the sample is obtained from. We compute the estimated quality $\hat{\rho}_i$ as

$$\hat{\rho}_i = \frac{1}{S_i} \sum_{s=1}^{S_i} \frac{c^i_{\alpha_t(i)}(t)}{\Gamma_{\alpha_t(i)}}.$$

such that $\hat{\rho}_i$ is an unbiased estimate of ρ_i (i.e., $\mathbb{E}_c[\hat{\rho}_i] = \rho_i$). By applying the Hoeffding's inequality we obtain a bound on the error of the estimated quality $\hat{\rho}_i$ for each ad i.

PROPOSITION 3.1. *For any ad $i \in \mathcal{N}$*

$$|\rho_i - \hat{\rho}_i| \leq \sqrt{\left(\sum_{s=1}^{S_i} \frac{1}{\Gamma^2_{\alpha_s(i)}} \right) \frac{1}{2S_i^2} \log \frac{n}{\delta}}, \tag{7}$$

with probability $1 - \delta$ (w.r.t. the click events).

It is easy to see that it is possible to define a sequence of explorative allocation strategies $\{\alpha_t\}_{t=1}^{\tau}$ such that $S_i = \lfloor K\tau/n \rfloor$, and (7) becomes[5]

$$|\rho_i - \hat{\rho}_i| \leq \sqrt{\left(\sum_{k=1}^{K} \frac{1}{\Gamma_k^2} \right) \frac{n}{2K^2\tau} \log \frac{n}{\delta}} := \eta, \tag{8}$$

After the exploration phase, an upper–confidence bound on each quality is computed as $\hat{\rho}_i^+ = \hat{\rho}_i + \eta$. From round τ on, the allocation $\hat{\alpha}$ ($\hat{\beta}$) simply sorts the advs according to their value $\hat{\rho}_i^+ v_i$ and allocates them in inverse order on each slot. Whenever the link

[5] From now on we drop the rounding and we use $\tau K/n$.

at slot k is clicked, the corresponding ad $\hat{\beta}(k)$ is charged with a payment

$$\tilde{p}_{\hat{\beta}(k)} = \frac{\sum_{l=k+1}^{K} (\Gamma_{l-1} - \Gamma_l) \max_i (\rho_i v_i; l)}{\Gamma_k \hat{\rho}^+_{\hat{\beta}(k)}}, \tag{9}$$

which results in an expected payment $\hat{p}_{\hat{\beta}(k)} = \tilde{p}_{\hat{\beta}(k)} \Gamma_k \rho_{\hat{\beta}(k)}$. The general form of the algorithm, which we refer to as A–VCG (Adaptive–VCG), is sketched in Figure 1.

We now move to the analysis of the performance of A–VCG in terms of the regret it cumulates through T rounds.

THEOREM 3.2. *Let us consider an auction with n advs, K slots, and T rounds. The auction have position–dependent externalities and cumulative discount factors $\{\Gamma_k\}_{k=1}^K$. For any parameter τ and δ, the A–VCG is always **truthful** and it achieves a regret*

$$R_T \leq V \Big(\sum_{k=1}^{K} \Gamma_k \Big) \Big(2(T-\tau)\eta + \tau + \delta T \Big). \tag{10}$$

By setting the parameters to

$$\delta = n^{1/3} (TK)^{-1/3}$$
$$\tau = 2^{1/3} T^{2/3} \Gamma_{\min}^{-2/3} K^{-1/3} n^{1/3} (\log (n^{2/3} (KT)^{1/3}))^{1/3},$$

where $\Gamma_{\min} = \min_k \Gamma_k \geq 0$, then the regret is

$$R_T \leq 18^{1/3} V T^{2/3} \Gamma_{\min}^{-2/3} K^{2/3} n^{1/3} (\log (n^2 KT))^{1/3}. \tag{11}$$

Remark 1 (The bound). Up to numerical constants and logarithmic factors, the previous bound is $R_T \leq \tilde{O}(T^{2/3} K^{2/3} n^{1/3})$. We first notice that A–VCG is a zero–regret algorithm since its per–round regret (R_T/T) decreases to 0 as $T^{-1/3}$, thus implying that it asymptotically achieves the same performance as the VCG. Furthermore, we notice that for $K = 1$ the bound reduces (up to constants) to the single–slot case analyzed in [Devanur and Kakade 2009]. Unlike the standard bound for multi–arm bandit algorithms, the regret scales as $\tilde{O}(T^{2/3})$ instead of $\tilde{O}(T^{1/2})$. As pointed out in [Devanur and Kakade 2009] and [Babaioff et al. 2008] this is the unavoidable price the bandit algorithm has to pay to be truthful. Finally, the dependence of the regret on n is sub-linear ($n^{1/3}$) and this allows to increase the number of advs without significantly worsening the regret.

Remark 2 (Distribution–free bound). The bound derived in Thm. 3.2 is a *distribution–free* (or worst–case) bound, since it holds for any set of advs (i.e., for any $\{\rho_i\}_{i \in \mathcal{N}}$ and $\{v_i\}_{i \in \mathcal{N}}$). This generality comes at the price that, as illustrated in other remarks and in the numerical simulations, the bound could be inaccurate for some specific sets of advs. On the other hand, distribution–dependent bounds (see e.g., the bounds of UCB [Auer et al. 2002]), where ρs and vs appear explicitly, would be more accurate in predicting the behavior of the algorithm. Nonetheless, they could not be used to optimize the parameters δ and τ, since they would then depend on unknown quantities (e.g., the qualities).

Remark 3 (Dependence on K). The most interesting property of the algorithm is its dependency on the number of slots K. According to the bound (11) the regret has a sublinear dependency $\tilde{O}(K^{2/3})$, meaning that whenever one slot is added to the auction, the performance of the algorithm does not significantly worsen. By analyzing

the difference between the payments of the VCG and A-VCG, we notice that during the exploration phase the regret is $O(\tau K)$ (e.g., if all K slots are clicked at each explorative round), while during the exploitation phase the error in estimating the qualities sum over all the K slots, thus suggesting a linear dependency on K for this phase as well. Nonetheless, as K increases, the number of samples available per each ad increases as $\tau K/n$, thus improving the accuracy of the quality estimates by $\tilde{O}(K^{-1/2})$ (see Proposition 3.1). As a result, as K increases, the exploration phase can be shortened (the optimal τ actually decreases as $K^{-1/3}$), thus reducing the regret during the exploration, and still have accurate enough estimations to control the regret of the exploitation phase.

Remark 4 (Parameters). The choice of parameters τ and δ reported in Thm. 3.2 is obtained by a rough minimization of the upper–bound (10) and can be computed by knowing the characteristics of the auction (number of rounds T, number of slots K, number of ads n, and cumulative discount factors Γ_k). Since they optimize an upper–bound, these values provide a good guess but they might not be the optimal parameterization for the problem at hand. Thus, in practice, the parameters could be optimized by searching the parameter space around the values suggested in Thm. 3.2.

PROOF. **(sketch)** *(Theorem 3.2)* The full proof is reported in the online appendix. Here we just focus on the per–round regret during the exploitation phase. According to the definition of payments in Sect. 3, at each round of the exploitation phase, the regret is the difference in the revenue, that is the difference in the expected payments. We bound the difference r between the payments in a slot k as

$$
\begin{aligned}
r &= \sum_{k=1}^{K} (p_{\beta^*(k)} - \hat{p}_{\hat{\beta}(k)}) \\
&= \sum_{k=1}^{K}\sum_{l=k}^{K} \Delta_l \left(\max_i(\rho_i v_i; l+1) - \frac{\max_i(\hat{\rho}_i^+ v_i; l+1)}{\hat{\rho}^+_{\hat{\beta}(k)}} \rho_{\hat{\beta}(k)} \right) \\
&\le \sum_{k=1}^{K}\sum_{l=k}^{K} \Delta_l \frac{\max_i(\hat{\rho}_i^+ v_i; l+1)}{\hat{\rho}^+_{\hat{\beta}(k)}} \left(\frac{\max_i(\rho_i v_i; l+1)}{\max_i(\hat{\rho}_i^+ v_i; l+1)} \hat{\rho}^+_{\hat{\beta}(k)} - \rho_{\hat{\beta}(k)} \right) \\
&= \sum_{k=1}^{K}\sum_{l=k}^{K} \frac{\Delta_l}{v^{-1}_{\hat{\beta}(k)}} \frac{\max_i(\hat{\rho}_i^+ v_i; l+1)}{\max_i(\hat{\rho}_i^+ v_i; k)} \left(\frac{\max_i(\rho_i v_i; l+1)}{\max_i(\hat{\rho}_i^+ v_i; l+1)} \hat{\rho}^+_{\hat{\beta}(k)} - \rho_{\hat{\beta}(k)} \right),
\end{aligned}
$$

where $\Delta_l = \Gamma_l - \Gamma_{l+1}$. By definition of the max operator, it follows that for any $l \ge k$, $\frac{\max_i(\hat{\rho}_i^+ v_i; l)}{\max_i(\hat{\rho}_i^+ v_i; k)} \le 1$. Using $v_{\hat{\beta}(k)} \le V$ (see Lemma A.1 in [A. Lazaric 2012]) and Proposition 3.1, it follows that

$$
\begin{aligned}
r &\le \sum_{k=1}^{K}\sum_{l=k}^{K} V \Delta_l (\hat{\rho}^+_{\hat{\beta}(k)} - \rho_{\hat{\beta}(k)}) \le V \sum_{k=1}^{K} (\hat{\rho}^+_{\hat{\beta}(k)} - \rho_{\hat{\beta}(k)}) \sum_{l=k}^{K} \Delta_l \\
&\le 2V\eta \sum_{k=1}^{K} \Gamma_k = V \Big(\sum_{k=1}^{K} \Gamma_k \Big) \sqrt{ \Big(\sum_{k=1}^{K} \frac{1}{\Gamma_k^2} \Big) \frac{2n}{K^2 \tau} \log \frac{n}{\delta} }
\end{aligned}
$$

with probability at least $1-\delta$. In order to get the final regret bound we need to consider the whole time horizon T and turn the bound into expectation. During the

first τ rounds A–VCG sets all the payments to 0 and the per–round regret is at most $V \sum_{k=1}^{K} \Gamma_k$, while in the remaining $T - \tau$ rounds the regret is bounded by r with probability $1 - \delta$. By adding all these terms, the statement follows. $\square$

(Comments to the proof). The proof uses relatively standard arguments to bound the regret of the exploitation phase. As discussed in Remark 2, the bound is distribution–free and some steps in the proof are conservative upper–bounds on quantities that might be smaller for specific auctions. For instance, the inverse dependency on the smallest cumulative discount factor $\Gamma_{\min}$ in the final bound could be a quite inaccurate upper–bound on the quantity $\sum_{k=1}^{K} 1/\Gamma_k^2$. In fact, the parameter τ itself could be optimized as a direct function of $\sum_{k=1}^{K} 1/\Gamma_k^2$, thus obtaining a more accurate tuning of the length of the exploration phase and a slightly tighter bound (in terms of constant terms). Furthermore, we notice that the step $\frac{\max_i(\hat{\rho}_i^+ v_i;l)}{\max_i(\hat{\rho}_i^+ v_i;k)} \leq 1$ is likely to become less accurate as the difference between l and k increases. For instance, if the qualities ρ_i are drawn from a uniform distribution in $(0, 1)$, as the number of slots increases this quantity reduces as well (on average) thus making the upper–bound by 1 less and less accurate. The accuracy of the proof and the corresponding bound are further studied in the simulations in Sect. 6.

4. POSITION/AD–DEPENDENT EXTERNALITIES

We now move to the general multi–slot model introduced in Sect. 2 where the discount factor of a slot k depends on the actual adv allocated on all the slots up to k. In this case the efficient allocation is $\alpha^* = \arg\max_\alpha \mathrm{SW}(\alpha)$. We notice that such maximization problem is often intractable since all the possible allocations of n ads over K slots should be tested in order to find the best one. The structure of the A–VCG algorithm (Figure 1) does not change. Nonetheless, the explorative allocations α_t have an impact on the discount $\Gamma_k(\alpha_t)$ and Proposition 3.1 becomes

$$|\rho_i - \hat{\rho}_i| \leq \sqrt{\left(\sum_{s=1}^{S_i} \frac{1}{\Gamma_{\alpha_s(i)}(\alpha_s)^2} \right) \frac{1}{2S_i^2} \log \frac{n}{\delta}}.$$

Similar to the previous section, we set $S_i = K\tau/n$ and we redefine η as

$$|\rho_i - \hat{\rho}_i| \leq \frac{1}{\Gamma_{\min}} \sqrt{\frac{n}{2K\tau} \log \frac{n}{\delta}} := \eta, \tag{12}$$

where $\Gamma_{\min} = \min_{\alpha,k} \Gamma_k(\alpha)$. We define the upper–confidence bound $\hat{\rho}_i^+ = \hat{\rho}_i + \eta$ and the estimated social welfare as

$$\widehat{\mathrm{SW}}(\alpha) = \widehat{\mathrm{SW}}(\beta) = \sum_{i=1}^{n} \Gamma_{\alpha(i)}(\alpha) \hat{\rho}_i^+ v_i = \sum_{k=1}^{n} \Gamma_k(\beta) \hat{\rho}_{\beta(k)}^+ v_{\beta(k)}.$$

The corresponding efficient allocation is denoted by $\hat{\alpha} = \arg\max_{\alpha \in \mathcal{A}} \widehat{\mathrm{SW}}(\alpha)$. Once the exploration phase is over, if ad $i \in \mathcal{N}$ is clicked, then the adv is charged

$$\tilde{p}_i = \frac{\widehat{\mathrm{SW}}(\hat{\alpha}_{-i}) - \widehat{\mathrm{SW}}_{-i}(\hat{\alpha})}{\Gamma_{\hat{\alpha}(i)} \hat{\rho}_i^+} \tag{13}$$

which corresponds to an expected payment $\hat{p}_i = \tilde{p}_i \Gamma_{\hat{\alpha}(i)} \rho_i$.

We are interested in bounding the regret of the A–VCG compared to the VCG.

THEOREM 4.1. *Let us consider an auction with n advs, K slots, and T rounds. The auction have position/ad–dependent externalities and cumulative discount factors*

$\{\Gamma_k(\alpha)\}_{k=1}^K$. *For any parameter τ and δ, the A–VCG is always* ***truthful*** *and it achieves a regret*

$$R_T \leq VK \left[(T-\tau) \left(\frac{3\sqrt{2}n}{\Gamma_{\min}\rho_{\min}} \sqrt{\frac{n}{K\tau} \log \frac{n}{\delta}} \right) + \tau + \delta T \right], \tag{14}$$

where $\rho_{\min} = \min_i \rho_i$. By setting the parameters to

$$\delta = n(TK)^{-1/3}$$
$$\tau = 18^{1/3} T^{2/3} \Gamma_{\min}^{-2/3} K^{-1/3} n (\log ((KT)^{1/3}))^{1/3},$$

the corresponding regret is

$$R_T \leq 6^{1/3} \frac{V}{\rho_{\min}} T^{2/3} \Gamma_{\min}^{-2/3} K^{2/3} n (\log (KT))^{1/3}. \tag{15}$$

Remark 1 (Differences with bound (11)). Up to constants and logarithmic factors, the previous distribution–free bound is $R_T \leq \tilde{O}(T^{2/3}K^{2/3}n)$. We first notice that moving from position– to position/ad–dependent externalities does not change the dependency of the regret on the number of rounds T. The main difference w.r.t. the bound in Thm. 3.2 is in the dependency on n and on the smallest quality $\rho_{\min}$. While the regret still scales as $K^{2/3}$, it has now a much worse dependency on the number of ads (from $n^{1/3}$ to n). We believe that it is mostly due to an intrinsic difficulty of the position/ad–dependent externalities. The intuition is that now in the computation of the payment for each ad i, the errors in the quality estimates cumulate through the slots (unlike the position–dependent case where they are scaled by $\Gamma_k - \Gamma_{k+1}$). Nonetheless, this cumulated error should impact only on a portion of the ads (i.e., those which are actually impressed according to the optimal and the estimated optimal allocations), while in the proof they are summed over all the advs. We conjecture that this additional n term is indeed a rough upper–bound on the number of slots K. If this were the case, we would obtain a regret $\tilde{O}(T^{2/3}K^{4/3}n^{1/3})$, where the dependency on the number of slots becomes super–linear. We postpone a more detailed analysis of this issue to Sect. 6. The other main difference is that now the regret has an inverse dependency on the smallest quality $\rho_{\min}$. Inspecting the proof, this dependency appears because the error of a quality estimation for an ad i might be amplified by the inverse of the quality itself ρ_i^{-1}. We investigate whether this dependency is an artifact of the proof or it is intrinsic in the algorithm in the numerical simulations reported in Sect. 6.

Remark 2 (Optimization of the parameter τ). We remark that whenever a guess about the value of $\rho_{\min}$ is available, it could be used to better tune τ by multiplying it by $\rho_{\min}^{-2/3}$, thus reducing the regret from $\tilde{O}(\rho_{\min}^{-1})$ to $\tilde{O}(\rho_{\min}^{-2/3})$.

Remark 3 (Externalities–dependent bound). We notice that the current bound does not reduce to (11) and thus it obviously over–estimates the dependency on K and n whenever the auction has position–dependent externalities. It is an interesting open question whether it is possible to derive an *auction–dependent* bound where the specific values of the discount factors $\gamma_k(\alpha)$ explicitly appear in the bound and that it reduces to (11) for position–dependent externalities.

(Comment to the proof). For the lack of space we do not report the proof, which can be found in the online appendix. While the proof of Thm. 3.2 could exploit the specific definition of the payments for position–dependent slots and it is a fairly standard

extension of [Devanur and Kakade 2009], in this case the proof is more complicated because of the dependency of the discount factors on the actual allocations and decomposes the regret of the exploitation phase in components due to the different allocations ($\hat{\alpha}$ instead of α^*) and the different qualities as well ($\hat{\rho}^+$ instead of ρ).

5. CONTEXTUAL MULTI–SLOT AUCTIONS

In real–world SSAs, the characteristics of the auctions (e.g., the quality) highly depend on contextual information such as the content of the webpage, the query submitted by the user, and her profile. In this section, we further generalize the auction with externalities to the case of contextual auctions. More precisely, we denote by $\mathcal{X}$ a subset of the Euclidean space $\mathbb{R}^s$ and we assume that $x \in \mathcal{X}$ summarizes all the contextual information necessary to define the auction. In particular, for each ad i, the quality is a function $\rho_i : \mathcal{X} \to [0,1]$, while we assume that the values v_i and the discount factors $\Gamma_k(\alpha)$ are independent from x.[6] When the functions ρ_i are known in advance, for each x we can still apply the VCG and charge adv a_i with a payment

$$p_i(x) = \mathrm{SW}(\alpha^*_{-i}; x) - \mathrm{SW}_{-i}(\alpha^*; x),$$

where $\mathrm{SW}(\alpha; x)$ is defined according to the qualities $\rho_i(x)$. The learning algorithm should now approximate each quality function over the whole domain $\mathcal{X}$. Although any regression algorithm could be employed to approximate ρ_i, here we consider a least squares approach for which performance bounds are available (see e.g., [Györfi et al. 2002]). We denote by $\phi(\cdot) = \big(\varphi_1(\cdot), \dots, \varphi_d(\cdot)\big)^\top$ a d-dimensional feature vector with features $\varphi_i : \mathcal{X} \to [0,1]$, and by $\mathcal{F} = \{f_w(\cdot) = \phi(\cdot)^\top w\}$ the linear space of functions spanned by the basis functions in ϕ. Similar to the previous settings, the algorithm first explores all the advs for τ rounds with an arbitrary exploration allocation α_t. At each round t, we assume a context x_t to be independently drawn from a stationary distribution μ over the context space $\mathcal{X}$. At the end of the exploration phase, each ad i has been impressed $S_i = K\tau/n$ times and we build the training set $\{x_s, c^i_{\alpha_s(i)}(s)\}_{s=1}^{S_i}$ where $c^i_{\alpha_s(i)}(s)$ is the click–event for ad i when displayed at slot $\alpha_s(i)$ in context x_s, and we compute the approximation

$$\hat{\rho}_i = f_{\hat{w}_i} = \arg\min_{f \in \mathcal{F}} \sum_{s=1}^{S_i} \Big(f(x_s) - \frac{c^i_{\alpha_s(i)}(s)}{\Gamma_{\alpha_s(i)}(\alpha_s)} \Big)^2 .$$

Since $\mathcal{F}$ is linear, we can easily compute the coefficient vector $\hat{w}_i$ in closed form. Let $\Phi_i = [\phi(x_s)^\top; \dots; \phi(x_{S_i})^\top]$ be the feature matrix corresponding to the training set and $c_i = (\frac{c^i_{\alpha_1(i)}(1)}{\Gamma_{\alpha_1(i)}(\alpha_1)}, \dots, \frac{c^i_{\alpha_1(S_i)}(S_i)}{\Gamma_{\alpha_{S_i}(i)}(\alpha_{S_i})})$ be the re-weighted vector of observations. Then we have

$$\hat{w}_i = (\Phi_i^\top \Phi_i)^{-1} \Phi_i^\top c_i. \tag{16}$$

During the exploitation phase, for any x, the A–VCG uses the quality $\hat{\rho}_i(x)$ to compute the allocation and define the payments. In particular, the estimated social welfare in a context x is defined as $\widehat{\mathrm{SW}}(\alpha; x) = \sum_{i=1}^n \Gamma_{\alpha(i)}(\alpha)\hat{\rho}_i(x)v_i$ and the expected payments become

$$\hat{p}_i(x) = \big(\widehat{\mathrm{SW}}(\hat{\alpha}_{-i}; x) - \widehat{\mathrm{SW}}_{-i}(\hat{\alpha}; x)\big)\frac{\rho_i(x)}{\hat{\rho}_i(x)}.$$

[6] The generalization to $v_i(x)$ and $\Gamma_k(\alpha; x)$ is straightforward.

Unlike the settings considered in the previous sections, we cannot expect to minimize the regret in each possible context $x \in \mathcal{X}$, thus we redefine the regret as the expectation w.r.t. the context distribution μ

$$R_{T,\mu} = T\mathbb{E}_{x\sim\mu}\Big[\sum_{i=1}^{n} p_i(x)\Big] - \sum_{t=1}^{T}\mathbb{E}_{x\sim\mu}\Big[\sum_{i=1}^{n} p_{it}(x)\Big],$$

where p_{it} is equal to 0 for the first $t \leq \tau$ explorative rounds and is equal to $\hat{p}_i$ during the exploitation phase.

In order to derive the regret bound, we need two technical assumptions (we further discuss them in the remarks).

ASSUMPTION 1. *The function space $\mathcal{F}$ contains all the quality functions ρ_i (i.e., the approximation error of $\mathcal{F}$ is 0), that is for any i*

$$\inf_{f\in\mathcal{F}} ||f - \rho_i||_\mu = 0.$$

ASSUMPTION 2. *The function space $\mathcal{F}$ is such that for any $f \in \mathcal{F}$, $\left\|1/f^2\right\|_\mu \leq \xi$.*

It is worth noting that here we no longer use an upper–confidence bound on ρ_i as before. In fact, it is not possible to build an upper–confidence bound for each context x since the accuracy of the approximated functions $\hat{\rho}_i$ can only be bounded in expectation w.r.t. the context distribution μ, as reported in the following lemma [Györfi et al. 2002].[7]

LEMMA 5.1. *Let $f_{\hat{w}_i}$ be computed as in (16) with $S_i = K\tau/n$ samples in a d-dimensional linear space $\mathcal{F}$, then for any $i \in \mathcal{N}$*

$$||f_{\hat{w}_i} - \rho_i||_\mu \leq \frac{64}{\Gamma_{\min}}\sqrt{\frac{(d+1)n}{K\tau}\log\big(\frac{324nTe^2}{\delta}\big)} := \chi \tag{17}$$

with probability $1-\delta$ (w.r.t. the random contexts and clicks), where $\Gamma_{\min} = \min_{\alpha,k}\Gamma_k(\alpha)$.

Given the two assumptions and Lemma 5.1, we can now derive the following regret bound.

THEOREM 5.2. *Let us consider a contextual auction on the domain $\mathcal{X}$ with n adv, K slots, and T rounds. Let μ be a distribution over $\mathcal{X}$. The auction have position/ad–dependent externalities and cumulative discount factors $\{\Gamma_k(\alpha)\}_{k=1}^{K}$. For any parameter τ and δ, the A–VCG is always truthful and it achieves a regret*

$$R_{T,\mu} \leq VK\left[6\xi(T-\tau)\chi + \tau + \delta T\right].$$

By setting the parameters to

$$\delta = n(TK)^{-1/3}$$
$$\tau = 24^{1/3}T^{2/3}\Gamma_{\min}^{-2/3}K^{-1/3}n(d+1)^{1/3}(\log\left(K^{1/3}T^{4/3}\right))^{1/3},$$

the corresponding regret is

$$R_{T,\mu} \leq 24^{1/3}V\xi T^{2/3}\Gamma_{\min}^{-2/3}K^{2/3}n(d+1)^{1/3}(\log\left(KT\right))^{1/3}.$$

Remark 1 (Bound). As we can notice, the bound obtained in the contextual case has exactly the same dependency on T, K, and n as in the regret in (15). The main difference is that two additional terms appear in the bound, the dimensionality of the

[7] We recall that the μ–weighted L_2–norm of a function f is defined as $||f||_\mu^2 = \mathbb{E}_{x\sim\mu}[f(x)^2]$.

space d and the lower–bound ξ on the functions in $\mathcal{F}$. It is interesting to notice that the regret grows as $\tilde{O}(d^{1/3})$ implying that the larger the number of features in $\mathcal{F}$, the worse the regret. This dependency is an immediate result of the fact that in (16) we learn a d-dimensional vector $\hat{w}_i$ and as d increases, the number of samples needed to have an accurate estimate of ρ_i increases as well, thus lengthening the exploration phase. Finally, the term ξ (see Assumption 2) plays a similar role as $\rho_{\min}^{-1}$ since it bounds the norm of the inverse of the functions in $\mathcal{F}$.

Remark 2 (Assumptions). Assumption 1 is the so–called *realizable* assumption, which implies that the functions to be approximated (the qualities ρ_i) belong to the function space $\mathcal{F}$. This assumption is reasonable whenever some prior knowledge about the ads is available and the features can be properly designed. Similarly, Assumption 2 is strictly related to the way the function space $\mathcal{F}$ is designed. In fact, it requires any function $f \in \mathcal{F}$ to be lower–bounded away from 0. This could be easily achieved by thresholding the prediction of each f. It is also worth noting that the two assumptions together imply that for any i, $\min_{x \in X} \rho_i(x) \geq \xi$, i.e., the advs have a quality which is at least ξ over the whole context space $\mathcal{X}$, thus suggesting that below a certain threshold the advs would not participate to the auction at all.

Remark 3 (Regression algorithm). Although in describing the algorithm we refer to least squares, any regression algorithm such as neural networks or logistic regression could be used. Nonetheless, in order to derive the regret bound, a specific replacement for Lemma 5.1 is needed.

6. NUMERICAL SIMULATIONS

In this section we report preliminary numerical simulations whose objective is to validate the theoretical results reported in the previous sections. In particular, we want to analyze how much the bounds accurately predict the dependency of the regret on the main characteristics of the auctions such as T, n, K, and $\rho_{\min}$. The advs are generated as follows. The qualities ρ_i are drawn from a uniform distribution in $[0.01, 0.1]$, while the values are randomly drawn from a uniform distribution on $[0, 1]$ ($V = 1$). Since the main objective is to test the accuracy of the bounds, we report the *relative regret*

$$\overline{R}_T = \frac{R_T}{B(T, K, n)},$$

where $B(T, K, n)$ is the value of the bound for the specific setting (i.e., position–dependent (11) and position/ad–dependent externalities (15)). We expect the relative regret to be always smaller than 1 (i.e., we expect B to be an actual upper–bound on the real regret R_T) and its value to be constant while changing one parameter and keeping all the others fixed. All the following results have been obtained by setting τ and δ as suggested by the bounds derived in the previous sections and by averaging over 100 independent runs. We leave the study of the contextual auction as a future work.

6.1. Position–Dependent Externalities

We first investigate auctions with position–dependent externalities. The discount factors are constant for all the positions (i.e., $\gamma_k = \gamma$) and are computed so that $\Gamma_1 = 1$ and $\Gamma_K = 0.8$, thus having $\Gamma_{\min} = 0.8$ in all the experiments. The left plot in Figure 2 shows the value of the relative regret $\overline{R}_T$ for different values of K and n when T increases. We notice that the three curves are completely flat and do not change as T increases. This suggests that the bound in Thm. 3.2 effectively predicts the dependency of the

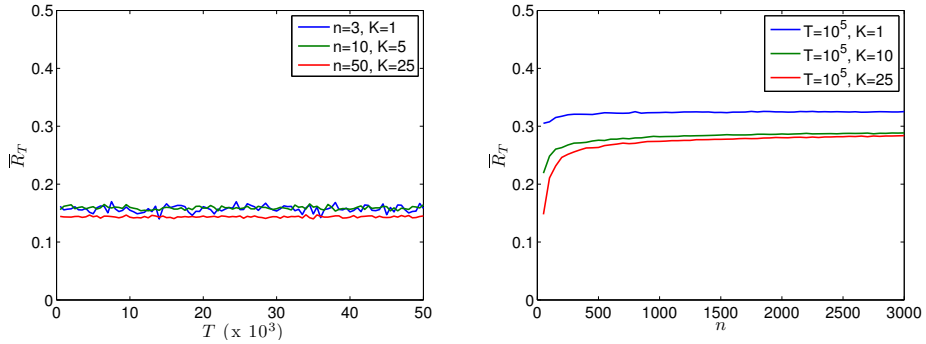

Fig. 2. Position–dependent externalities. Dependency of the relative regret on T, n.

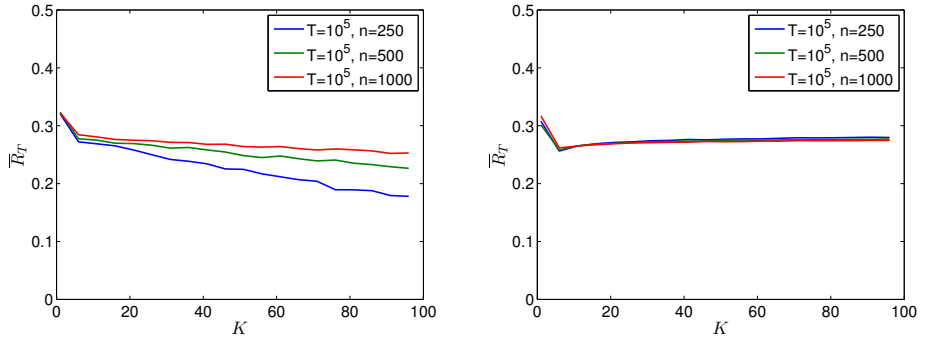

Fig. 3. Position–dependent externalities. Dependency of the relative regret on K for two different choice of the the qualities ρ.

regret R_T w.r.t. the number of rounds T of the auction, i.e., $\tilde{O}(T^{2/3})$. The second plot in Figure 2 shows the dependency on the number of ads n. In this case we notice that it is relatively accurate as n increases but there is a transitory effect for smaller values of n where the regret grows faster than predicted by the bound (although $B(T, K, n)$ is still an upper–bound to R_T). Finally, the first plot in Figure 3 on the left suggests that the dependency on K in the bound of Thm. 3.2 is over–estimated, since the relative regret $\overline{R}_T$ decreases as K increases. As discussed in the comment to the proof in Sect. 3 this might be explained by the over–estimation of the term $\frac{\max_i(\hat{\rho}_i^+ v_i; l)}{\max_i(\hat{\rho}_i^+ v_i; k)}$ in the proof. In fact, this term is likely to decrease as K increases. In order to test this intuition, we ran an additional experiment where ρs are such that $\rho_1 = 0.1$, $\rho_2 = 0.095$, and all the others are equal to 0.09. As a result, the ratio between the qualities ρ_i is fixed (on average) and does not change with K. We report the result in the second plot of Figure 3. For different choices of n the ratio $\overline{R}_T$ is constant, implying that in this case the bound accurately predicts the behavior of R_T. This result confirms Remark 2 in Section 5, suggesting that $K^{2/3}$ is the right dependency of R_T on the number of slots in the worst–case w.r.t. the qualities, while it could be better for some specific ρs.

6.2. Position/Ad–Dependent Externalities

We now study the accuracy of the bound derived in Thm. 4.1 where the regret R_T displays a linear dependency on n and an inverse dependency on the smallest quality $\rho_{\min}$. The relative regret $\overline{R}_T$ is now defined as R_T/B with B the bound (15). In the first plot of Figure 4 we report $\overline{R}_T$ as T increases. As it can be noticed, the bound accurately predict the behavior of the regret w.r.t. T as in the case of position–dependent

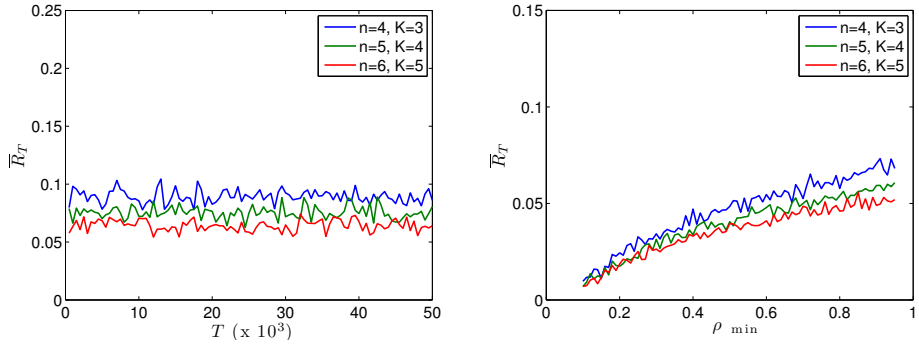

Fig. 4. Dependency on T and $\rho_{\min}$ in auctions with position/ad–dependent externalities.

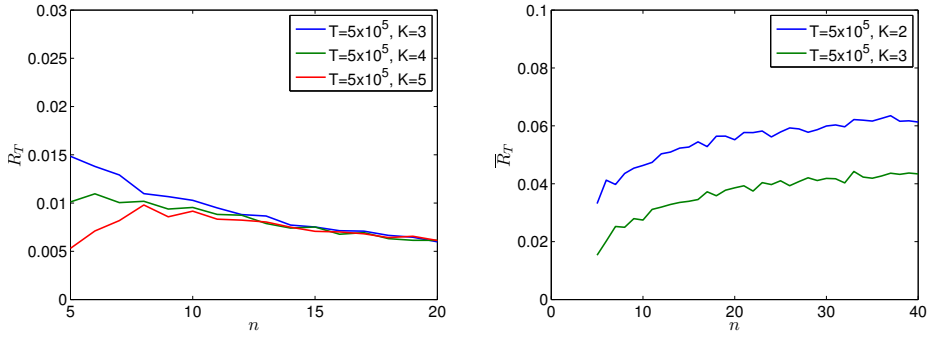

Fig. 5. Dependency of the relative regret $\overline{R}_T$ on n. In the first plot $\overline{R}_T$ is defined as R_T/B with B the bound (14) and $\tau = \tilde{O}(n^{1/3})$, while in the second plot we use B as in the bound (11).

externalities. In the second plot of Figure 4 we report $\overline{R}_T$ as we change $\rho_{\min}$. According to the bound in (15) the regret should decrease as $\rho_{\min}$ increases (i.e., $R_T \leq \tilde{O}(\rho_{\min}^{-1})$) but it is clear from the plot that R_T has a much smaller dependency on $\rho_{\min}$, if any[8]. Finally, we study the dependency on n. As conjectured in Sect. 4, we believe that (15) is over–estimating the actual dependency on n. We set $\tau = \tilde{O}(n^{1/3})$ which, according to the bound (14), should lead to a regret $R_T = \tilde{O}(n^{4/3})$. As reported in the left plot of Figure 5 this dependency is largely over–estimated and as n increases $\overline{R}_T$ actually decreases. Thus we tried to study whether the bound (10) for position–dependent could be used also in this setting. Since $\tau = \tilde{O}(n^{1/3})$ is the optimal choice for position–dependent externalities, we expect to obtain $R_T = \tilde{O}(n^{1/3})$. Thus we recompute the relative $\overline{R}_T$ by dividing R_T by the bound (11) of Theorem 3.2 when $\tau = \tilde{O}(n^{1/3})$. As it can be noticed in Figure 5, $\overline{R}_T$ now slightly increases and then it tends to flat as n increases. This resembles the same behavior as in the second plot of Figure 2, suggesting that R_T might indeed have a similar dependency $n^{1/3}$ w.r.t. the number of ads. Nonetheless, this does not exclude the possibility that the dependency on the number of slots K might get worse (from $K^{2/3}$ to $K^{4/3}$) as conjectured in Remark 1 of Theorem 4.1. We believe that further experiments and a more detailed theoretical analysis are needed. In fact, as commented in Remark 2 of Sect. 3, all the upper–bounds are distribution–free and consider worst–case auctions (in terms of qualities, values, and discount factors). As a

[8]From this experiment is not clear whether $\overline{R}_T = \tilde{O}(\rho_{\min})$, thus implying that R_T does not depend on $\rho_{\min}$ at all, or $\overline{R}_T$ is sublinear in $\rho_{\min}$, which would correspond to a dependency $R_T = \tilde{O}(\rho_{\min}^{-\alpha})$ with $\alpha < 1$.

result, it is difficult to claim whether the linear dependency n could be indeed obtained for some specific auctions or it is due to a rough maximization of the number of slots as conjectured in Remark 1 of Thm. 4.1. Finally, we mention that we do not report results on K since the complexity of finding the optimal allocation α^* becomes intractable as for $K > 5$.

7. CONCLUSIONS AND FUTURE WORK

Multi–arm bandit is an effective framework to study the quality estimation problem in sponsored search auctions. In this paper, we extended the truthful exploration–exploitation mechanism defined in [Devanur and Kakade 2009] to the general case of contextual multi–slot auctions with position/ad–dependent externalities. The upper–bounds on the revenue regret show an explicit dependency on the number of advs n and slots K in the auction and they have been largely confirmed by numerical simulations. This work open several questions:

Estimation of the cumulative discount factors. Through all the paper we assumed that A–VCG receives as input the (exact) cumulative discount factors and only the qualities $\{\rho_i\}_i$ are estimated. This assumption is reasonable in the case of position–dependent externalities where the coefficients $\{\Gamma_k\}_k$ can be easily estimated using historical data on any other similar auctions (independently from which advs actually participated in it). On the other hand, for position/ad–dependent externalities the coefficients $\{\Gamma_k\}_k$ depend on the actual advs and they might not be easier to estimate than the qualities themselves. Nonetheless, the algorithm and the analysis illustrated in the paper could both be extended to the general case where $\{\Gamma_k\}_k$ are estimated at the same time as the qualities. The main difficulty is that each ad should be displayed on all the slots in order to distinguish between the probability of click which strictly depend on the ad (i.e., ρ_i) and the probability due to the specific slot and allocation wherein it is impressed. This would result in a significant lengthening of the exploration phase and a corresponding worsening of the regret.

Truthfulness. As shown in [Devanur and Kakade 2009] and [Babaioff et al. 2008], truthfulness has a major impact in the achievable regret. It is interesting to understand whether relaxed notions of truthfulness (e.g., with high probability, ϵ–truthfulness) or moving to a characterization of *ex post* implementation (see e.g., [Langford et al. 2010]) could lead to recover the standard multi–arm bandit $\tilde{O}(\sqrt{T})$ regret (see also [Slivkins 2011] for similar issues in the case of monotone multi–armed bandit allocations).

Auction–dependent bound. As discussed in Remark 2 of Thm. 3.2 and Remark 3 of Thm. 4.1, it would be interesting to derive auction–dependent bounds where the qualities, values, and discount factors, explicitly appear.

Lower bounds. Although a lower bound on the regret as a function of T is available for the single–slot case, it is crucial to derive lower bounds for the multi–slot case with an explicit dependency on n and K.

Approximated auctions. As pointed out in Sect. 4, the computation of the efficient allocation α^* is intractable. In [Kempe and Mahdian 2008] a quasi–polynomial ϵ–approximation of efficient allocation is derived. An interesting direction of future work is to derive a regret analysis for such an approximated mechanism.

Experiments. The numerical simulations reported here are simple experiments to study the accuracy of the bounds we derived in the paper. It is important to test the A–VCG on real datasets, in particular in the contextual case.

REFERENCES

A. LAZARIC, N. GATTI, F. T. 2012. A truthful learning mechanism for contextual

multi–slot sponsored search auctions with externalities. Tech. Rep. al-00662549, version 1.

AGGARWAL, G., FELDMAN, J., MUTHUKRISHNAN, S., AND PÁL, M. 2008. Sponsored search auctions with markovian users. In *Proceedings of the 4th International Workshop on Internet and Network Economics (WINE'08)*. 621–628.

AUER, P., CESA-BIANCHI, N., AND FISCHER, P. 2002. Finite-time analysis of the multiarmed bandit problem. *Machine Learning Journal (MLJ) 47*, 235–256.

BABAIOFF, M., SHARMA, Y., AND SLIVKINS, A. 2008. Characterizing truthful multi-armed bandit mechanisms. *CoRR abs/0812.2291*.

DEVANUR, N. R. AND KAKADE, S. M. 2009. The price of truthfulness for pay-per-click auctions. In *Proceedings of the 10th ACM Conference on Electronic Commerce (EC'09)*. 99–106.

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.

EVEN-DAR, E., MANNOR, S., AND MANSOUR, Y. 2006. Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *Journal of Machine Learning Research (JMLR) 7*, 1079–1105.

GONEN, R. AND PAVLOV, E. 2007a. An adaptive sponsored search mechanism δ-gain truthful in valuation, time and budget. In *Proceedings of the 3rd International Workshop on Internet and Network Economics (WINE'07)*. 341–346.

GONEN, R. AND PAVLOV, E. 2007b. An incentive-compatible multi-armed bandit mechanism. In *Proceedings of the twenty-sixth annual ACM symposium on Principles of distributed computing (PODC'07)*. 362–363.

GYÖRFI, L., KOHLER, M., KRZYŻAK, A., AND WALK, H. 2002. *A distribution-free theory of nonparametric regression*. Springer-Verlag, New York.

IAB. 2010. IAB internet advertising revenue report. 2010 first half-year results.

KEMPE, D. AND MAHDIAN, M. 2008. A cascade model for externalities in sponsored search. In *Proceedings of the 4th International Workshop on Internet and Network Economics (WINE'08)*. 585–596.

LANGFORD, J., LI, L., VOROBEYCHIK, Y., AND WORTMAN, J. 2010. Maintaining equilibria during exploration in sponsored search auctions. *Algorithmica 58*, 990–1021.

MAS-COLELL, A., WHINSTON, M., GREEN, J., AND DE CIÈNCIES ECONÒMIQUES I EMPRESARIALS, U. P. F. F. 1995. *Microeconomic theory*. Oxford University Press New York.

NARAHARI, Y., GARG, D., NARAYANAM, R., AND PRAKASH, H. 2009. *Game Theoretic Problems in Network Economics and Mechanism Design Solutions*. Springer.

NAZERZADEH, H., SABERI, A., AND VOHRA, R. 2008. Dynamic cost-per-action mechanisms and applications to online advertising. In *Proceeding of the 17th international conference on World Wide Web (WWW'08)*. 179–188.

PANDEY, S. AND OLSTON, C. 2006. Handling Advertisements of Unknown Quality in Search Advertising. In *Proceedings of the Conference on Neural Information Processing Systems (NIPS'06)*. 1065–1072.

ROBBINS, H. 1952. Some aspects of the sequential design of experiments. *Bulletin of the AMS 58*, 527–535.

SLIVKINS, A. 2011. Monotone multi-armed bandit allocations. In *Proceedings of the 24th Annual Conference on Learning Theory (COLT'11) - Open Problems Track*. Vol. 19. 829–834.

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