Horizon-Independent Optimal Pricing in Repeated Auctions with Truthful and Strategic Buyers

Alexey Drutsa
Yandex, 16, Leo Tolstoy St.
Moscow, Russia
adrutsa@yandex.ru

ABSTRACT

We study revenue optimization learning algorithms for repeated posted-price auctions where a seller interacts with a (truthful or strategic) buyer that holds a fixed valuation. We focus on a practical situation in which the seller does not know in advance the number of played rounds (the time horizon) and has thus to use a horizon-independent pricing. First, we consider straightforward modifications of previously best known algorithms and show that these horizon-independent modifications have worser or even linear regret bounds. Second, we provide a thorough theoretical analysis of some broad families of consistent algorithms and show that there does not exist a no-regret horizon-independent algorithm in those families. Finally, we introduce a novel deterministic pricing algorithm that, on the one hand, is independent of the time horizon T and, on the other hand, has an optimal strategic regret upper bound in $O(\log \log T)$. This result closes the logarithmic gap between the previously best known upper and lower bounds on strategic regret.

Keywords: Repeated auctions; revenue optimization; horizon-independent pricing; strategic regret; reserve price; posted-price auction

1. INTRODUCTION

Revenue optimization in online advertising is one of the most important development direction in large and modern Internet companies (such as search engines [48, 3, 53, 26, 16], social networks [1], real-time ad exchanges [25, 12], etc.). Auctions play a vital and central role in this area [13, 42]: the most applicable ones are second-price [26, 35, 45], generalized second-price (GSP) [48, 34, 46, 16], and Vickrey-Clarke-Groves (VCG) [49, 50] auctions, where revenue is mainly controlled by means of setting proper reserve prices [40, 32]. This reflects in the recent explosion in the number of published studies on a more algorithmic approaches to optimize revenue of auctions, including machine learned reserve prices [15, 26, 5, 28, 35, 52, 46, 37, 36, 43, 45, 44]. A large number of online auctions run by, e.g., ad exchanges involve only a single bidder [5, 37], and, in this case, a second-price auction with reserve is equivalent to a *posted-price auction* [31] where the seller sets a reserve price for a good (e.g., an advertisement space) and the buyer decides whether to accept or reject it (i.e., to bid above or below the price).

We study a scenario when the seller repeatedly interacts through a posted-price mechanism with the *same* buyer that holds a *fixed* private valuation for a good. The seller's goal is to maximize his revenue over a finite number of rounds T (*the time horizon*), that is generally reduced to *regret*[1] minimization, and the seller thus seeks for a *no-regret* pricing algorithm, i.e., with a sublinear regret on T [5, 37, 6, 38, 17]. In a simple setting, when the buyer behaves *truthfully* (i.e., *myopically*: accepts an offered price if and only if it is no larger than his valuation), the seller can apply the fast search algorithm [31] that admits an optimal truthful regret upper bound in $O(\log \log T)$. In a more sophisticated setting, when the buyer behaves *strategically* [5, 37] seeking to maximize his cumulative γ-discounted surplus over T rounds[2], the seller can apply the algorithm PFS [37] that has a nearly optimal strategic regret upper bound in $O(\log T \log \log T)$.

The main weakness of the existing algorithms [31, 5, 37] is their strong dependence on the time horizon (namely, the algorithms' parameters being set independently of T imply linear regrets), because, in practice, it is very natural that the seller does not know in advance the number of rounds T that the buyer wants to interact with him. Hence, *in the current work, we focus on horizon-independent pricing algorithms* that could be used by the seller in this situation.

On the one hand, to the best of our knowledge, no existing studies on our scenario with fixed private valuation considered this aspect (the studies [31, 5, 37] on this scenario neither addressed horizon independence as well). On the other hand, there is the state-of-the-art technique, known as "doubling trick" [15, 27, 20] and "squaring trick" [4, 54, 33, 20], that constructs a horizon-independent algorithm from a horizon-dependent one and that was earlier applied to algorithms of other scenarios (e.g., a stochastic buyer valuation [15, 20] and a buyer's pricing algorithm [27]). We adapt this technique (introducing *the "exponentiating trick"*) to the existing algorithms [31, 37] of our scenario (see Sec. 4) and show that the modified variants admit similar upper regret

WWW 2017, April 3–7, 2017, Perth, Australia.
ACM 978-1-4503-4913-0/17/04.
http://dx.doi.org/10.1145/3038912.3052700

[1] In this scenario, the regret is the difference between the revenue that would have been earned by offering the buyer's valuation and the seller's revenue (see Sec. 3.1 and [31, 37]).

[2] This setting is motivated by the insight (supported by empirical observations [23]): the knowledge that a seller use a revenue optimization algorithm may incite the buyer to mislead the seller and boost the buyer's surplus [5, 37].

bounds as the original ones (e.g., a non-optimal bound in the case of PFS). Moreover, the upgraded algorithms regularly "reset" their learning and do not thus exploit the historical buyer behavior before a reset round (that may unnecessarily increase the regret). However, since the buyer holds a fixed valuation, a good online algorithm that learns on the buyer decisions made at the past rounds should probably work consistently [37]: after an acceptance, set only no lower prices (*right consistency*) and, after a rejection, set only no higher prices (*left consistency*) than the currently offered one.

Therefore, the primary research goal of our work is to construct horizon-independent online learning (reinforcement learning) algorithms for setting prices that admit an optimal regret bound in both truthful and strategic settings of our scenario and are as much consistent as possible. Our study is developed in a step-by-step manner. In each buyer behavior setting, first, we identify the key reasons why algorithms may admit a linear regret and formally establish this reasons via a theorem on a regret lower bound in a certain class of algorithms. Second, we propose an algorithm (beyond this class), which avoids the identified causes of a linear regret, and provide theoretical guarantees for its optimality.

In the truthful setting (Sec. 5.1), we show that a linear regret is caused either by non-density of the algorithm prices or by a non-decaying fraction of price rejections (w.r.t. the growth of T) along some buyer strategies. Hence, we propose the consistent algorithm FES that: (a) infinitely conducts an exploration of the buyer's valuation (thus, the algorithm prices are dense) and, (b) when the buyer makes a rejection, exploits the last accepted price with a rate that growths double exponentially w.r.t. the number of rejections (thus, the algorithm never faces a non-decaying fraction of rejections).

In the strategic setting (Sec. 5.2), additionally to the issues of the truthful one, we indicate that the linear regret can be caused by the ability of the buyer to exploit the left consistency and force thus a consistent pricing algorithm to offer prices lower than $v - \varepsilon$ (v is the valuation, $\varepsilon > 0$) in order to get the maximal surplus for him and, hence, a linear regret for the seller. Hence, we seek for a no-regret algorithm beyond the class of consistent ones (namely, we relax the left consistency condition) and propose the right-consistent algorithm PRRFES that: additionally to the options (a) and (b), (c) applies penalization repeats of a rejected price forcing the buyer to lie less (similarly to [37]) and (d) regularly revises rejected prices. We show that with a proper selection of the algorithm's parameter, if a price is rejected due to a lie of the buyer, then this price will be accepted in some future round (i.e., the strategic buyer has no incentive to infinitely receive a price lower than $v - \varepsilon$, for a fixed $\varepsilon > 0$).

The most surprising fact in our study is that, while seeking for a horizon-independent algorithm, we built the algorithm PRRFES that has a tight strategic regret upper bound in $\Theta(\log\log T)$. This, in fact, closes the previously open research question on the existence of an algorithm (even among not only horizon-independent!) with a more favorable regret bound than $O(\log T \log\log T)$ (achieved by PFS [37]), since the known strategic regret lower bound is $\Omega(\log\log T)$.

To sum up, our paper focuses on the problem, which meets the *present and emerging Internet companies' needs*: to maximize revenue of frequently used online advertising mechanisms. Specifically, the major contributions of our study are fundamental and include:

- Novel optimal horizon-independent pricing algorithms FES and PRRFES for repeated posted-price auctions with truthful and strategic buyers, respectively, that thus outperform the existing algorithms upgraded by the state-of-the-art "doubling"/"squaring" tricks.

- Closing of the logarithmic gap between the previously best known upper and lower bounds on strategic regret by constructing an algorithm with $O(\log\log T)$ regret.

- Linear lower bound on the strategic (truthful) regret of any horizon-independent pricing algorithm that is regular weakly (strongly, respectively) consistent.

The rest of the paper is organized as follows. In Sec. 2, the related work on auctions is discussed. In Sec. 3, we state the problem and give background on pricing algorithms. The "exponentiating trick" is presented in Sec. 4. Sec. 5 contains our main findings: the study of horizon-independent algorithms with consistent proprieties that includes theoretical guarantees. In Sec. 6, the conclusions are provided.

2. RELATED WORK

A large body of studies on online advertising auctions lies in the field of game theory [32]: most of them focused on characterizing different aspects of equilibria, and recent ones was devoted (but not limited) to: position auctions [48, 49, 50, 16], different second-price auction extensions [3, 14], efficiency [2], mechanism expressiveness [22], competition across auction platforms [8], buyer budget [1], experimental analysis [42, 47, 41], etc.

Studies on revenue optimization were devoted to both the seller revenue solely [53, 26] and different sort of trade-offs either between several auction stakeholders [25, 24, 10] or between auction properties (like expressivity, simplicity [39], and revenue monotonicity [24]). The optimization problem was generally reduced to a selection of proper quality scores for advertisements (for auctions with several advertisers [53, 26]) or reserve prices for buyers (e.g., for VCG [40], GSP [34], and others [25, 43]). The latter ones, in such setups, usually depend on distributions of buyer bids or valuations and was in turn estimated by machine learning techniques [26, 46, 43], while alternative approaches learned reserve prices directly [35, 36, 45]. In contrast to these works, we use an online deterministic learning approach for repeated auctions.

Revenue optimization for repeated auctions was mainly concentrated on algorithmic reserve prices, that are updated in online fashion over time, and was also known as dynamic pricing, see [21], where an extensive survey on this field could be found. Dynamic pricing was studied: under game-theoretic view (MFE [29, 12], budget constraints [12, 11], strategic buyer behavior [18], dynamic mechanisms [7], etc.); as bandit problems [4, 54, 33] (e.g., UCB-like pricing [9], bandit feedback models [51]); from the buyer side (valuation learning [29, 51], competition between buyers and optimal bidding [28, 51], interaction with several sellers [27], etc.); from the seller side against several buyers [15, 52, 30, 44]; and a single buyer with stochastic valuation (truthful [31, 19] and strategic buyers [5, 6, 38, 38, 17], feature-based pricing [6, 20], limited supply [9], etc.). The most relevant part of these works on online learning is the state-of-the-art technique (known as "doubling" [15, 27, 20] and "squaring" [4, 54, 33, 20] tricks) that build a horizon-independent algorithm from a horizon-dependent one and, to the best of our knowledge, was never studied for algorithms of our fixed-valuation scenario. We adapt this approach to our case by proposing the "exponentiating trick" in Sec. 4. Overall, the most relevant studies to ours are [31, 5, 37], where our scenario with a fixed private valuation is considered and whose algorithms will be discussed in more details in Sec. 3.3. In contrast to these works, we, first, study algorithms that are independent of the time horizon T and, second, propose one of them that has a tight strategic regret bound in $\Theta(\log\log T)$.

3. FRAMEWORK

3.1 Setup of repeated posted-price auctions

We consider the following scenario of *repeated posted-price auctions* [5, 37]. A good (e.g., an advertisement space) is repeatedly offered for sale by a seller to a *single* buyer over T rounds (*the time horizon*). The buyer holds a private *fixed* valuation $v \in [0,1]$ for that good (which is unknown to the seller). At each round $t \in \{1,\dots,T\}$, a price p_t is offered by the seller, and an allocation decision $a_t \in \{0,1\}$ is made by the buyer: $a_t = 1$, when the buyer accepts to buy a currently offered good at that price, 0, otherwise. Thus, the seller applies a *(pricing) algorithm* $\mathcal{A}$ that sets prices $\{p_t\}_{t=1}^T$ in response to buyer decisions $\mathbf{a} = \{a_t\}_{t=1}^T$ referred to as a *(buyer) strategy*. We consider the deterministic online learning case when the price p_t at a round $t \in \{1,\dots,T\}$ can depend only on the buyer's actions during the previous rounds $\mathbf{a}_{1:t-1}$[3] and the horizon T. Hence, given $\mathcal{A}$, a strategy $\mathbf{a}$ uniquely defines the corresponding price sequence $\{p_t\}_{t=1}^T$.

Given a time horizon T, a pricing algorithm $\mathcal{A}$, and a buyer strategy $\mathbf{a} = \{a_t\}_{t=1}^T$, the seller's total revenue is $\sum_{t=1}^T a_t p_t$, where the price sequence $\{p_t\}_{t=1}^T$ corresponds to the strategy $\mathbf{a}$. This revenue is usually compared to the revenue that would have been earned by offering the buyer's valuation v if it was known in advance to the seller [31, 5, 37]. This leads to the definition of the *regret* of the algorithm $\mathcal{A}$ that faced a buyer with the valuation $v \in [0,1]$ following the (buyer) strategy $\mathbf{a}$ over T rounds as

$$\mathrm{Reg}(T,\mathcal{A},v,\mathbf{a}) := \sum_{t=1}^{T}(v - a_t p_t).$$

Truthful setting. Let us assume that the buyer does not exploit the seller's behavior (he is *myopic*) or, alternatively, as in [31], one can assume that the seller interacts with a different buyer at each round. In this case, the buyer accepts a price whenever it is no larger than his valuation v, i.e., his strategy is $\mathbf{a}^{\mathrm{Truth}}(\mathcal{A},v)$ defined by $a_t^{\mathrm{Truth}} := \mathbb{I}_{\{p_t \le v\}}$[4]. Thus, we define *the truthful regret* of the algorithm $\mathcal{A}$ that faced a truthful buyer with valuation $v \in [0,1]$ over T rounds as

$$\mathrm{TReg}(T,\mathcal{A},v) := \mathrm{Reg}\big(T,\mathcal{A},v,\mathbf{a}^{\mathrm{Truth}}(\mathcal{A},v)\big).$$

Strategic setting. Following a standard assumption in mechanism design that matches the practice in ad exchanges [37], the pricing algorithm $\mathcal{A}$, used by the seller, *is announced to the buyer in advance.* In this case, the buyer can act strategically against this algorithm: we assume that the buyer follows the optimal strategy $\mathbf{a}^{\mathrm{Opt}}(T,\mathcal{A},v,\gamma)$ that maximizes the buyer's γ-discounted surplus [5], $\gamma \in (0,1]$,

$$\mathrm{Sur}_\gamma(T,\mathcal{A},v,\mathbf{a}) := \sum_{t=1}^{T}\gamma^{t-1}a_t(v - p_t),$$

i.e., $\mathbf{a}^{\mathrm{Opt}}(T,\mathcal{A},v,\gamma) := \mathrm{argmax}_{\mathbf{a}}\,\mathrm{Sur}_\gamma(T,\mathcal{A},v,\mathbf{a})$. Thus, we define *the strategic regret* of the algorithm $\mathcal{A}$ that faced a strategic buyer with valuation $v \in [0,1]$ over T rounds as

$$\mathrm{SReg}(T,\mathcal{A},v,\gamma) := \mathrm{Reg}\big(T,\mathcal{A},v,\mathbf{a}^{\mathrm{Opt}}(T,\mathcal{A},v,\gamma)\big).$$

Hence, we consider a two-player non-zero sum repeated game with incomplete information and unlimited supply, introduced by Amin et al. [5] and considered in [37]: the buyer seeks to maximize his surplus, while the seller's objective is to minimize his strategic regret (i.e., maximize his revenue). Note that the discount factor is presented only in the buyer's objective (not in the seller's one), which is motivated by the observation that, in important real-world markets (including online advertising), sellers are far more willing to wait for revenue than buyers are willing to wait for goods [5, 37].

For each setting, following [31, 5, 6, 37, 38], we are interested in algorithms that attain $o(T)$ strategic (truthful) regret (i.e., the averaged regret goes to zero as $T \to \infty$) for the worst-case valuation $v \in [0,1]$, i.e., we say that an algorithm $\mathcal{A}$ is *no-regret* when $\sup_{v\in[0,1]}\mathrm{Reg}(T,\mathcal{A},v,\mathbf{a}) = o(T)$ for $\mathbf{a} = \mathbf{a}^{\mathrm{Opt}}$ ($\mathbf{a}^{\mathrm{Truth}}$ resp.). Namely, we seek for algorithms that have the lowest possible strategic (truthful) regret upper bound of the form $O(f(T))$ and treat their optimality in terms of $f(T)$ with the slowest growth as $T \to \infty$ (the averaged regret has thus the best rate of convergence to zero).

3.2 Notations and auxiliary definitions

For a fixed $T \in \mathbb{N}$, a deterministic pricing algorithm $\mathcal{A}$ can be associated with a complete binary tree $\mathfrak{T}(\mathcal{A})$ of depth $T-1$ [31, 37]. Each node $\mathfrak{n} \in \mathfrak{T}(\mathcal{A})$[5] is labeled with the price $p^{\mathfrak{n}}$ offered by $\mathcal{A}$. The right and left children of $\mathfrak{n}$ are denoted by $\mathfrak{r}(\mathfrak{n})$ and $\mathfrak{l}(\mathfrak{n})$ respectively. The left (right) subtrees rooted at node $\mathfrak{l}(\mathfrak{n})$ ($\mathfrak{r}(\mathfrak{n})$ resp.) are denoted by $\mathfrak{L}(\mathfrak{n})$ ($\mathfrak{R}(\mathfrak{n})$ resp.).[6].

So, the algorithm works as follows: it starts at the root $\mathfrak{n}_1$ of the tree $\mathfrak{T}(\mathcal{A})$ by offering the first price $p^{\mathfrak{n}_1}$ to the buyer; at each step $t < T$, if a price $p^{\mathfrak{n}}, \mathfrak{n} \in \mathfrak{T}(\mathcal{A})$, is accepted, the algorithm moves to the right node $\mathfrak{r}(\mathfrak{n})$ and offers the price $p^{\mathfrak{r}(\mathfrak{n})}$; in the case of the rejection, it moves to the left node $\mathfrak{l}(\mathfrak{n})$ and offers the price $p^{\mathfrak{l}(\mathfrak{n})}$; this process repeats until reaching a leaf. The round at which the price of a node $\mathfrak{n} \in \mathfrak{T}(\mathcal{A})$ is offered is denoted by $t^{\mathfrak{n}}$ (it is equal to the node's depth $+1$). Note that each node $\mathfrak{n} \in \mathfrak{T}(\mathcal{A})$ uniquely determines the buyer decisions up to the round $t^{\mathfrak{n}} - 1$. Thus, each buyer strategy $\mathbf{a}_{1:t}$ is bijectively mapped to a t-length path in the tree $\mathfrak{T}(\mathcal{A})$ that starts from the root and goes to a t-depth node (and the strategy prices are the ones that are in the nodes lying along this path).

We define, for a pricing tree $\mathfrak{T}$, the set of its prices $\wp(\mathfrak{T}) := \{p^{\mathfrak{n}} \mid \mathfrak{n} \in \mathfrak{T}\}$ and denote by $\wp(\mathcal{A}) := \wp(\mathfrak{T}(\mathcal{A}))$ all prices that can be offered by an algorithm $\mathcal{A}$. We say that two complete trees $\mathfrak{T}_1$ and $\mathfrak{T}_2$ of depth d_1 and d_2, resp., are *price equivalent* and write $\mathfrak{T}_1 \cong \mathfrak{T}_2$ if the trees have the same node

[3] We use a notation for a part of a strategy $\mathbf{a}_{t_1:t_2} = \{a_t\}_{t=t_1}^{t_2}$.

[4] $\mathbb{I}_B$ is the indicator: $\mathbb{I}_B = 1$, when B holds, and 0, otherwise.

[5] For simplicity, if $\mathfrak{n}$ is a node of a tree $\mathfrak{T}$, we write $\mathfrak{n} \in \mathfrak{T}$.

[6] Note that, in order to simplify notations in our definitions and proofs in Sec. 5, we use slightly different to [31, 37] notions of the algorithm tree (depth $T-1$ instead of T) and the right/left subtrees (rooted at $\mathfrak{r}(\mathfrak{n})/\mathfrak{l}(\mathfrak{n})$ instead of $\mathfrak{n}$).

labeling when we naturally match the nodes between the trees (starting from the roots) up to the depth $\leq \min\{d_1, d_2\}$ (i.e., following the same strategy in both trees, the buyer receives the same sequence of prices).

3.3 Background on pricing algorithms

Since the buyer holds a fixed valuation, we could expect that a smart online pricing algorithm should work consistently: after an acceptance (a rejection), set only no lower (no higher, resp.) prices than the offered one. Formally,

Definition 1. An algorithm $\mathcal{A}$ is said to be *consistent* [37] ($\mathcal{A}$ in the class $\mathbf{C}$) if, for any node $\mathfrak{n} \in \mathfrak{T}(\mathcal{A})$,

$$p^{\mathfrak{m}} \geq p^{\mathfrak{n}} \quad \forall \mathfrak{m} \in \mathfrak{R}(\mathfrak{n}) \quad \text{and} \quad p^{\mathfrak{m}} \leq p^{\mathfrak{n}} \quad \forall \mathfrak{m} \in \mathfrak{L}(\mathfrak{n}).$$

The key idea behind a consistent algorithm $\mathcal{A}$ is clear: it explores the valuation domain $[0, 1]$ by means of a feasible search interval $[q, q']$ (initialized by $[0, 1]$) targeted to locate the valuation v. At each round t, $\mathcal{A}$ offers a price $p_t \in [q, q']$ and, depending on the buyer's decision, reduces the interval to the right subinterval $[p_t, q']$ (by $q := p_t$) or the left one $[q, p_t]$ (by $q' := p_t$); at any moment, q is thus always the last accepted price or 0, while q' is the last rejected price or 1. The most known example of a consistent algorithm is the binary search.

The consistency represents a quite reasonable property, when the buyer is truthful, because a reported buyer decision correctly locates v in $[0, 1]$. For this setting, Kleinberg et al. [31] proposed the *Fast Search (FS) algorithm* that keeps track of a feasible interval $[q, q']$ initialized to $[0, 1]$ and an increment parameter ϵ initialized to $1/2$. The algorithm works in phases within *the exploration stage*: within each phase, it offers prices $q + \epsilon$, $q + 2\epsilon, \ldots$ until a price is rejected. If a price $q+k\epsilon$ is rejected, then a new phase starts with the new interval $[q, q'] := [q + (k-1)\epsilon, q + k\epsilon]$ and the new increment parameter $\epsilon := \epsilon^2$. This process continues until $q' - q < 1/T$, and the price q is offered all the remaining rounds (*the exploitation stage*). The authors proved that the truthful regret of this algorithm is upper bounded by $O(\log_2 \log_2 T)$. They also showed that the truthful regret of any pricing algorithm is lower bounded by $\Omega(\log_2 \log_2 T)$ [31]. Hence, the FS algorithm is optimal in terms of the seller truthful regret.

In the strategic setting, the buyer, incited by surplus maximization, may mislead the seller's consistent algorithm [6, 37]. Amin et al. [5] showed that, for $\gamma = 1$, any algorithm has a linear strategic regret by proving a necessary condition for no-regret pricing: the buyer horizon $T_\gamma = \sum_{t=1}^{T} \gamma^{t-1}$ should be $o(T)$. For this case of $\gamma \in (0, 1)$, there were proposed two no-regret algorithms. The first one is *the Monotone algorithm* [5]: it offers prices $p_t = \beta^{t-1}, \beta \in (0, 1)$, until the one of them is accepted, then this price is offered all the remaining rounds. The second one is *the Penalized Fast Search (PFS) algorithm* [37]: it follows the pricing of FS algorithm, but, when a price p_t is rejected by the buyer, the seller offers this price for the next $r-1$ additional rounds (penalization), $r \in \mathbb{N}$; if all of them are rejected, PFS continues the FS pricing; if the buyer accepts the price at a penalization round, then the seller apply the same pricing as if the buyer accepts the price p_t first time at the round t (for $r = 1$, PFS matches FS). For our further needs, we give the following formal definition related to the penalization rounds:

Definition 2. Nodes $\mathfrak{n}_1, \ldots, \mathfrak{n}_r \in \mathfrak{T}(\mathcal{A})$ are said to be a (r-length) penalization sequence if $\mathfrak{n}_{i+1} = \mathfrak{l}(\mathfrak{n}_i)$,

$$p^{\mathfrak{n}_{i+1}} = p^{\mathfrak{n}_i}, \quad \text{and} \quad \mathfrak{R}(\mathfrak{n}_{i+1}) \cong \mathfrak{R}(\mathfrak{n}_i), \quad i = 1, \ldots, r-1.$$

It is easy to see that a strategic buyer either accepts the price at the first node or rejects this price in all of them.

The Monotone's strategic regret has tight bound in $\Theta(T^{1/2})$, when $\beta = T^{1/2}/(1 + T^{1/2})$ [37]. The PFS's strategic regret is upper bounded by $O(\log_2 T \log_2 \log_2 T)$, when selecting a proper number of penalization rounds to force the buyer lie less, namely, $r = \lceil \mathrm{argmin}_{r' \geq 1} r' + \frac{\gamma_0^{r'} T}{(1-\gamma_0)(1-\gamma_0^{r'})} \rceil$, for $1/2 < \gamma < \gamma_0 < 1$; and by $O(\log_2 \log_2 T)$, when $r = 1$, for $\gamma \in (0, 1/2]$. The known lower bound of the strategic regret for any pricing algorithm is $\Omega(\log_2 \log_2 T)$, the same as in the truthful case.

Overall, in the truthful setting, there exists an optimal algorithm, while, in the strategic setting, the existence of an algorithm with the strategic regret bounded in $O(\log_2 \log_2 T)$ has remained open for $\gamma \in (1/2, 1)$ (PFS is nearly optimal: there is the logarithmic gap between its upper and lower bounds). We close this research question by proposing our algorithm PRRFES and proving its optimality in Sec. 5.2.

3.4 Horizon-independent pricing

Note that, in the previous subsections, we talk about algorithms that may depend on the time horizon T (they are called non-uniform deterministic pricing algorithms as well [31]). We can indicate it in an algorithm's notation as $\mathcal{A}(T)$, and the trees $\mathfrak{T}(\mathcal{A}(T))$ may not comprise each other for different T (i.e., the labels (prices) in the trees may be different in corresponding nodes of the same depth).

However, in practice, e.g., of ad exchanges, it is very natural that the seller does not know in advance the number of rounds T that the buyer wants to interact with him. Hence, in the current study, we focus on pricing algorithms that do not depend on the a priori knowledge of the time horizon T and could be used by the seller in this situation. We refer to an algorithm $\mathcal{A}$ of this sort as a *horizon-independent* one, also referred to as an uniform deterministic pricing algorithm [31], for which there is a single infinite tree $\mathfrak{T}$ whose first $T-1$ levels comprise $\mathfrak{T}(\mathcal{A}(T))$ for each $T \in \mathbb{N}$. Therefore, since we mainly study algorithms of this sort, for simplicity of notations in those place where it will not lead to a misunderstanding, we assume that the tree $\mathfrak{T}(\mathcal{A})$ of a horizon-independent algorithm $\mathcal{A}$ is infinite, can admit infinite descending paths (i.e., infinite buyer strategies) with infinite corresponding price sequences $\{p_t\}_{t=1}^{\infty}$. Note that the game remains finite, and we still consider the buyer that maximizes his surplus over finite T rounds (the case of infinite horizon for the surplus is discussed in Sec. 5.3).

All previously known algorithms from Sec. 3.3 (FS, Monotone, and PFS) have to know the horizon T in advance in order to be no-regret (their parameters depend on T, e.g., FS has the exploration termination rule $q' - q < 1/T$). Note that straightforward ways to make them be horizon independent will not succeed in a no-regret pricing: e.g., if the exploration stage in FS/PFS stops independently of T, see Corollary 1 and if the exploration stage is not stopped, see Theorem 2. In the following section, we adapt to the algorithms from Sec. 3.3 the state-of-the-art technique that upgrade an algorithm to horizon-independent one, and show that the upgraded variants admit the same upper regret bounds as the original ones. Since they are not optimal in the strategic setting, we proceed to seek for horizon-independent optimal algorithms with some consistent properties in Sec. 5.

4. EXPONENTIATING TRICK

In the studies on stochastic-valuation scenarios and bandit problems, there is the state-of-the-art technique (known as "doubling" [15, 27, 20] and "squaring" [4, 54, 33, 20] tricks) that makes a horizon-independent algorithm from a horizon-dependent one and which we adapt to our case by proposing "exponentiating trick". Namely, given a horizon-dependent algorithm $\mathcal{A}(T)$, the idea is to partition time $\mathbb{T} := \mathbb{N}$ into epochs $\{\mathbb{T}_i\}_{i\in\mathbb{N}}$, $\mathbb{T}_i = \{\sum_{j=1}^{i-1} T_j + 1, \ldots, \sum_{j=1}^{i} T_j\}$, of increasing lengths $|\mathbb{T}_i| = T_i, i \in \mathbb{N}$. Let a function $h : \mathbb{N} \to \mathbb{N}$, referred to as *the magnification rate*, define the epoch lengths as follows: $T_i = h(T_{i-1})$ with some $T_1 \in \mathbb{N}$. We apply the algorithm $\mathcal{A}(T_i)$ at each epoch $\mathbb{T}_i, i \in \mathbb{N}$, and obtain a horizon-independent algorithm $\widetilde{A}_h$.

If $h(n) = 2n$ (doubling, $T_i = 2^{i-1}T_1$) and $h(n) = n^2$ (squaring, $T_i = T_1^{2^{i-1}}$), then the technique is referred to as "doubling trick" [15, 27, 20] and "squaring trick" [4, 54, 33, 20], respectively. If a regret of the original pricing $\mathcal{A}$ is upper bounded in $O(T^\alpha)$ (or $O(\log^\alpha T)$), then $\widetilde{A}_h$ with "doubling" (or "squaring" resp.) trick will satisfy the same regret upper bounds (with a larger constant hidden in $O(\cdot)$). So, "doubling trick" is fine for Monotone, but FS and PFS have double logarithmic growth in their upper bounds. So, if we apply the "doubling" ("squaring") trick to the algorithms FS and PFS, the obtained modifications will have less favorable upper bounds w.r.t. the ones of FS and PFS: they will increase by a factor in $O(\log_2 T)$ ($O(\log_2\log_2 T)$ resp.), to see this, please, follow the proof of Th. 1. Thus, a direct application of the state-of-the-art technique to these algorithms of our fixed-valuation scenario does not give us the best regret upper bounds. Therefore, we propose *the exponentiating magnification rate* $h_E(n) = n^{\log_2 n}$, and, thus, *the "exponentiating trick"*, for which (when $T_1 = 4$) we have the following growth of epoch lengths: $\log_2\log_2 T_i = 2^{i-1}, i \in \mathbb{N}$.

THEOREM 1. *Given $\gamma_0 \in (1/2, 1)$, let $\widetilde{\mathrm{FS}}_{h_E}$ and $\widetilde{\mathrm{PFS}}_{h_E}$ be the FS and PFS algorithms, respectively, upgraded by the "exponentiating trick" (i.e., epochs built on the magnification rate $h_E(n) = n^{\log_2 n}$), then, for any valuation $v \in [0, 1]$,*

$$\mathrm{TReg}(T, \widetilde{\mathrm{FS}}_{h_E}, v) = O(\log_2\log_2 T),$$
$$\mathrm{SReg}(T, \widetilde{\mathrm{PFS}}_{h_E}, v, \gamma) = O(\log_2\log_2 T), \gamma \in (0, 1/2], \quad \textit{and}$$
$$\mathrm{SReg}(T, \widetilde{\mathrm{PFS}}_{h_S}, v, \gamma) = O(\log_2 T \log_2\log_2 T), \gamma \in (1/2, \gamma_0).$$

PROOF SKETCH. Due to the space constraints and since our trick is quite similar to "squaring" one [33], we provide only the proof sketch and only for the second equation (the others are similar). Let $R(T, \mathcal{A}) := \mathrm{SReg}(T, \mathcal{A}, v, \gamma)$, then $R(T, \mathrm{PFS}(T)) \le c\log_2\log_2 T, c > 0$ (see Sec. 3.3). Let $\hat{T}_i = \sum_{j=1}^{i} T_j$, where $\log_2\log_2 T_i = 2^{i-1}, i \in \mathbb{N}$, by def. of h_E, then, given $T \in \mathbb{T}_k$ (i.e., the horizon observes k epochs), one has $T_{k-1} \le \hat{T}_{k-1} < T \le \hat{T}_k$ and $k - 2 < \log_2\log_2\log_2 T$. In each epoch $\mathbb{T}_i, i \le k$, the strategic buyer's behavior in response to $\widetilde{\mathrm{PFS}}_{h_E}$ is the same as in response to $\mathrm{PFS}(T_i)$ (since the pricing during $\mathbb{T}_i$ does not depend on the one during $\mathbb{T}_j, j < i$). Moreover, for the case of the last epoch, one can show that the strategic buyer's behavior over $T - \hat{T}_{k-1}$ rounds in response to $\mathrm{PFS}(T_k)$ results in $R(T - \hat{T}_{k-1}, \mathrm{PFS}(T_k)) \le R(T_k, \mathrm{PFS}(T_k))$ (see the proof of [37, Th. 1] and slightly improve it). Therefore, one can estimate the regret:

$$R(T, \widetilde{\mathrm{PFS}}_{h_E}) \le \sum_{j=1}^{k} R(T_j, \mathrm{PFS}(T_j)) = c\sum_{j=1}^{k} 2^{j-1}$$
$$\le c \cdot 2(2^{k-1} - 1) < 4c\log_2\log_2 T,$$

for any $T > T_1 = 4$. □

Thus, we obtained a horizon-independent algorithm (i.e., $\widetilde{\mathrm{FS}}_{h_E}$) with an optimal truthful regret upper bound, and the one (i.e., $\widetilde{\mathrm{PFS}}_{h_E}$) with a nearly optimal strategic regret upper bound (similar to PFS). Overall, first, we has not obtained an algorithm with an optimal upper bound on strategic regret. Second, modifications based on the technique are not consistent algorithms, since they do no exploit the information from previous epochs, that may unnecessarily increase the regret (e.g., see the proof of Th. 1: the constant in $O(\cdot)$ is $\times 4$ w.r.t. the one of the non-modified algorithm). Therefore, we move to study consistent horizon-independent algorithms, that may have more favorable properties.

5. CONSISTENT ALGORITHMS

Several types of algorithm consistency will be of particular interest in our further study. We introduce them (beside the class **C** from Definition 1) in the following definitions. We start from the subclass of consistent algorithms that each time offer a new price (never exploit previous ones):

Definition 3. An algorithm $\mathcal{A}$ is said to be *strongly consistent* ($\mathcal{A}$ in the class **SC**) if, for any node $\mathfrak{n} \in \mathfrak{T}(\mathcal{A})$,

$$p^{\mathfrak{m}} > p^{\mathfrak{n}} \ \ \forall\mathfrak{m} \in \mathfrak{R}(\mathfrak{n}) \quad \text{and} \quad p^{\mathfrak{m}} < p^{\mathfrak{n}} \ \ \forall\mathfrak{m} \in \mathfrak{L}(\mathfrak{n}).$$

Definition 4. An algorithm $\mathcal{A}$ is said to be *weakly consistent* ($\mathcal{A}$ in the class **WC**) if, for any node $\mathfrak{n} \in \mathfrak{T}(\mathcal{A})$,

$$\text{when } \mathfrak{r}(\mathfrak{n}) \text{ s.t. } p^{\mathfrak{r}(\mathfrak{n})} \ne p^{\mathfrak{n}}, \quad p^{\mathfrak{m}} \ge p^{\mathfrak{n}} \ \ \forall\mathfrak{m} \in \mathfrak{R}(\mathfrak{n}) \quad \text{and,}$$
$$\text{when } \mathfrak{l}(\mathfrak{n}) \text{ s.t. } p^{\mathfrak{l}(\mathfrak{n})} \ne p^{\mathfrak{n}}, \quad p^{\mathfrak{m}} \le p^{\mathfrak{n}} \ \ \forall\mathfrak{m} \in \mathfrak{L}(\mathfrak{n}).$$

Weakly consistent algorithms are similar to consistent ones, but they are additionally able to offer the same price p several times before making a final decision on which of the subintervals $[q, p]$ or $[p, q']$ continue (see Sec. 3.3). This class is introduced to comprise the algorithm PFS [37], that is not consistent for $r > 1$ due to the penalization rounds (see Def. 2). However, the class **WC** is too large. Hence, we consider its subclass that can also wait with the subinterval decision, but the pricing will be the same no matter when a decision is made (it also contains the algorithm PFS).

Definition 5. A weakly consistent algorithm $\mathcal{A}$ is said to be *regular* ($\mathcal{A}$ in the class **RWC**) if, for any node $\mathfrak{n} \in \mathfrak{T}(\mathcal{A})$:

1. when $p^{\mathfrak{l}(\mathfrak{n})} = p^{\mathfrak{n}} = p^{\mathfrak{r}(\mathfrak{n})}$,
$$\left[p^{\mathfrak{m}} = p^{\mathfrak{n}} \ \ \forall\mathfrak{m} \in \mathfrak{R}(\mathfrak{l}(\mathfrak{n})) \cup \mathfrak{L}(\mathfrak{r}(\mathfrak{n}))\right] \text{ or } \left[\mathfrak{L}(\mathfrak{n}) \cong \mathfrak{R}(\mathfrak{n})\right];$$
2. when $p^{\mathfrak{l}(\mathfrak{n})} = p^{\mathfrak{n}} \ne p^{\mathfrak{r}(\mathfrak{n})}$,
$$\left[p^{\mathfrak{m}} = p^{\mathfrak{n}} \ \ \forall\mathfrak{m} \in \mathfrak{R}(\mathfrak{l}(\mathfrak{n}))\right] \quad \text{or} \quad \left[\mathfrak{R}(\mathfrak{l}(\mathfrak{n})) \cong \mathfrak{R}(\mathfrak{n})\right];$$
3. when $p^{\mathfrak{l}(\mathfrak{n})} \ne p^{\mathfrak{n}} = p^{\mathfrak{r}(\mathfrak{n})}$,
$$\left[p^{\mathfrak{m}} = p^{\mathfrak{n}} \ \ \forall\mathfrak{m} \in \mathfrak{L}(\mathfrak{r}(\mathfrak{n}))\right] \quad \text{or} \quad \left[\mathfrak{L}(\mathfrak{r}(\mathfrak{n})) \cong \mathfrak{L}(\mathfrak{n})\right].$$

Definition 6. An algorithm $\mathcal{A}$ is said to be *right-consistent* ($\mathcal{A}$ in the class $\mathbf{C_R}$) if, for any $\mathfrak{n} \in \mathfrak{T}(\mathcal{A})$, $p^{\mathfrak{m}} \geq p^{\mathfrak{n}} \forall \mathfrak{m} \in \mathfrak{R}(\mathfrak{n})$.

Right-consistent algorithms never offer a price lower than the last accepted one, but may offer a price larger than a rejected one (in contrast to consistent algorithms). Overall, it is easy to see that the following relations between the defined classes of consistency (the sets of algorithms) holds:

$$\mathbf{SC} \subset \mathbf{C} \subset \mathbf{RWC} \subset \mathbf{WC} \qquad \text{and} \qquad \mathbf{C} \subset \mathbf{C_R}.$$

Before analyzing pricing algorithms for truthful and strategic settings, we consider a common necessary condition to be a no-regret algorithm. A buyer strategy $\mathbf{a}$ is said to be *locally non-losing* (w.r.t. v and $\mathcal{A}$) if prices greater than v are never accepted[7] (i.e., $a_t = 1$ implies $p_t \leq v$).

Definition 7. An algorithm $\mathcal{A}$ is said to be *dense* if the set of its prices $\wp(\mathcal{A})$ is dense in $[0, 1]$ (i.e., $\overline{\wp(\mathcal{A})} = [0, 1]$).

Lemma 1. *If a horizon-independent pricing algorithm $\mathcal{A}$ is not dense then there exists a valuation $v \in [0, 1]$ s.t., for any locally non-losing strategy $\mathbf{a}$, $\mathrm{Reg}(T, \mathcal{A}, v, \mathbf{a}) = \Omega(T)$.*

Proof. Since the prices $\wp(\mathcal{A})$ are not dense in $[0, 1]$, there exist $\varepsilon > 0$ and $v \in (0, 1)$ s.t. $(v - \varepsilon, v + \varepsilon) \subset [0, 1] \setminus \overline{\wp(\mathcal{A})}$. Hence, for any $T > 0$ and for any locally non-losing strategy $\mathbf{a}$ with the corresponding sequence of prices $\{p_t\}_{t=1}^{T}$, we have $p_t < v - \varepsilon$ for all $t = 1, \ldots, T$ s.t. $a_t = 1$, and, thus,

$$\mathrm{Reg}(T, \mathcal{A}, v, \mathbf{a}) > \sum_{t:a_t=0} v + \sum_{t:a_t=1} \big(v - (v - \varepsilon)\big) \geq T\varepsilon.$$

This lower bound is $\Omega(T)$ since ε is independent of T. □

Note that, first, the truthful buyer's strategy $\mathbf{a}^{\mathrm{Truth}}$ is locally non-losing one by its definition. Second, in the case of $\mathcal{A} \in \mathbf{C_R}$, the optimal buyer strategy $\mathbf{a}^{\mathrm{Opt}}$ is locally non-losing one as well (by right-consistency, once accepting a price $p_t > v$, the buyer will receive $p_{t'} > v$, $\forall t' > t$, and will thus suffer from a negative surplus after the round t). The same holds for the case of $\mathcal{A} \in \mathbf{RWC}$: the buyer has no incentive to accept a price $p_t > v$, since he will receive either no lower prices, or the same price as if he rejected the price at the t-th round. Hence, we immediately get the following.

Corollary 1. *For any non-dense horizon-independent algorithm $\mathcal{A}$, there exists a valuation $v \in [0, 1]$ such that* $\mathrm{TReg}(T, \mathcal{A}, v) = \Omega(T)$. *Moreover, if $\mathcal{A}$ is right- or regular weakly consistent, then* $\mathrm{SReg}(T, \mathcal{A}, v, \gamma) = \Omega(T)$ $\forall \gamma \in (0, 1]$.

5.1 Truthful setting

In this subsection, for the truthful setting, we show, first, that there does not exist a no-regret horizon-independent algorithm in the class $\mathbf{SC}$ (Theorem 2). Second, we present our no-regret horizon-independent algorithm FES from the class $\mathbf{C}$ and prove its optimality (Theorem 3).

Proposition 1. *Let $\mathcal{A}$ be a dense horizon-independent consistent pricing algorithm, then the sequence of prices of any buyer strategy converges.*

Algorithm 1 Pseudo-code of the FES pricing algorithm

```
Input: $g : \mathbb{Z}_+ \to \mathbb{Z}_+$
Initialize: $q := 0$, $p := 1/2$, $l := 0$, $k := 1$
while the buyer plays do
    Offer the price $p$ to the buyer
    if the buyer accepts the price then
        $q := p$
    else
        Offer the price $q$ to the buyer for $g(l)$ rounds
        if the buyer rejects one of the prices then
            Offer the price $q$ until the buyer stops playing
        end if
        $l := l + 1, \quad k := 0$
    end if
    if $k < 2^{2^{l-1}}$ then
        $p := q + 2^{-2^{l}}, \quad k := k + 1$
    else
        Offer the price $p$ until the buyer stops playing
    end if
end while
```

Proof. Let us consider any strategy $\mathbf{a}$ with the corresponding sequence of prices $\{p_t\}_{t=1}^{\infty}$. We denote

$$\underline{p} = \liminf_{t\to\infty} p_t \quad \text{and} \quad \overline{p} = \limsup_{t\to\infty} p_t.$$

If $\underline{p} < \overline{p}$, then let us show that $(\underline{p}, \overline{p})$ does not contain any price of the algorithm. First, for the strategy $\mathbf{a}$, $p_t \notin (\underline{p}, \overline{p})$ $\forall t \in \mathbb{N}$. Indeed, if there exists $t_0 \in \mathbb{N}$ such that $p_{t_0} \in (\underline{p}, \overline{p})$, then, in the case of $a_{t_0} = 0$, $p_t \leq p_{t_0}$ $\forall t > t_0$ (due to the consistency) and, hence, $\overline{p} \leq p_{t_0}$, but it is a contradiction to the assumption $p_{t_0} < \overline{p}$. The case $a_{t_0} = 1$ could be considered in a similar way.

Second, for any strategy $\mathbf{a}'$ with prices $\{p'_t\}_{t=1}^{\infty}$, if $\mathbf{a}' \succ \mathbf{a}$, i.e., there exists $t_0 \in \mathbb{N}$ such that $a'_{t_0} > a_{t_0}$ and $a'_t = a_t \, \forall t < t_0$. Hence, $p'_t = p_t \, \forall t \leq t_0$, $a'_{t_0} = 1$, and $a_{t_0} = 0$, that implies

$$p'_{t'} \geq p'_{t_0} = p_{t_0} \geq p_t \qquad \text{for any } t', t \geq t_0,$$

where we used the consistency of the algorithm $\mathcal{A}$. One thus has $p'_{t'} \geq \overline{p} \, \forall t' \geq t_0$, and $p'_{t'} = p_{t'} \notin (\underline{p}, \overline{p}) \, \forall t' < t_0$. In a similar way, for any strategy $\mathbf{a}'' \prec \mathbf{a}$ with prices $\{p''_t\}_{t=1}^{\infty}$, we have $p''_t \leq \underline{p} \, \forall t \geq t_0$, and $p'_t = p_t \notin (\underline{p}, \overline{p}) \, \forall t < t_0$, for some $t_0 \in \mathbb{N}$. Therefore, $(\underline{p}, \overline{p})$ contains no algorithm's price from $\wp(\mathcal{A})$ (i.e., $(\underline{p}, \overline{p}) \subset [0, 1] \setminus \overline{\wp(\mathcal{A})}$), the algorithm $\mathcal{A}$ is thus not dense, and we obtain a contradiction. Otherwise, $\underline{p} = \overline{p}$, and this is equivalent to the existence of the limit $\lim_{t\to\infty} p_t$. □

Theorem 2. *For any horizon-independent strongly consistent pricing algorithm $\mathcal{A}$, there exists a valuation $v \in [0, 1]$ s.t.* $\mathrm{TReg}(T, \mathcal{A}, v) = \Omega(T)$.

Proof. If the algorithm $\mathcal{A}$ is not dense, then the theorem holds due to Corollary 1. For a dense algorithm, we consider a strategy $\mathbf{a}$ defined by $a_t := \mathbb{I}_{\{t \mod 2 = 0\}}, t \in \mathbb{N}$, (i.e., it alternates a rejection and an acceptance) with its corresponding price sequence $\{p_t\}_{t=1}^{\infty}$. By Proposition 1, there exists the limit $p = \lim_{t\to\infty} p_t$. For $t = 2s - 1, s \in \mathbb{N}$, i.e., the reject rounds ($a_t = 0$), any further price $p_{t'} < p_t \forall t' > t$, and, hence, the limit $p \leq p_t$. Moreover, if $p = p_t$, then, by *the strong consistency* of the algorithm $\mathcal{A}$, $p \leq p_{t+2} < p_t = p$, which is a contradiction. Therefore, the limit $p < p_t$. Similarly, for $t = 2s, s \in \mathbb{N}$, i.e., the accept rounds ($a_t = 1$),

[7] Note that the optimal strategy of a strategic buyer may not satisfy this property: it is easy to imagine an algorithm that offers the price 1 at the first round and, if it is accepted, offers the price 0 all remaining rounds.

one can show that the limit $p > p_t$. Thus, we shown that $a_t = \mathbb{I}_{\{p_t \le p\}} \equiv \mathbb{I}_{\{p_t < p\}}$ (since $p \neq p_t \, \forall t \in \mathbb{N}$).

Let us take the price limit as the buyer valuation $v := p$, then, $\mathbf{a}$ is the truthful strategy of the buyer with this valuation, and this truthful buyer will thus reject a price in a half of played rounds. Hence, $\mathrm{TReg}(T, \mathcal{A}, v) \ge v\lceil T/2 \rceil$. □

Note that, in the proof, one can replace the strategy $\mathbf{a}$ by any sequence with a non-decaying fraction of rejections as $T \to \infty$ and get a bunch of valuations v that yield a linear truthful regret. This theorem shows us: a no-regret pricing that explores prices all rounds (e.g., FS without the stop-criteria $\epsilon < 1/T$) does not exist.

FES algorithm. We take the idea of the algorithm FS and improve it to avoid the causes of a linear regret showed in Lemma 1 (Corollary 1) and Theorem 2: we (a) conduct exploration infinitely and (b) inject an exploitation with a growing rate after each rejection. Formally, our *Fast Exploiting Search pricing algorithm* (*FES*) is consistent and works against a truthful strategy in phases initialized by the phase index $l := 0$, the last accepted price before the current phase $q_0 := 0$, the iteration parameter $\epsilon_0 := 1/2$, and the number of offers $N_0 := 2$; at each phase $l \in \mathbb{Z}_+$, it sequentially offers prices $p_{l,k} := q_l + k\epsilon_l, k = 1, .., N_l$, (exploration), where

$$\epsilon_l := \epsilon_{l-1}^2 = 2^{-2^l},\ N_l := \epsilon_{l-1}/\epsilon_l = \epsilon_{l-1}^{-1} = 2^{2^{l-1}},\ l \in \mathbb{N}; \quad (1)$$

if a price $p_{l,k}$ with $k = K_l + 1 \ge 1$ is rejected, (1) it offers the price p_{l,K_l} for $g(l)$ rounds (exploitation) and (2) FES goes to the next phase by setting $q_{l+1} := p_{l,K_l}$ and $l := l+1$. The pseudo-code of FES is presented in Alg. 1, which describes the full algorithm even in the case of facing a non-truthful strategy. Note that the lines 10 and 17 in Algorithm 1 are never reached by any truthful buyer, but are introduced in the pseudo-code in order to formally satisfy the consistent conditions (for the case when the algorithm faces a non-truthful strategy): thus, FES is in the class $\mathbf{C}$.

The function $g : \mathbb{Z}_+ \to \mathbb{Z}_+$ is the parameter of our algorithm, which is referred to as *the exploitation rate*. We set it as

$$g(l) = 2^{2^l}, \quad l \in \mathbb{Z}_+, \quad (2)$$

which growths double exponentially w.r.t. the number of rejections. This allows us properly avoid the main cause of linear regret in Th. 2 (a non-decaying fraction of rejections along a truthful strategy) and prove the following theorem.

THEOREM 3. *Let $\mathcal{A}$ be the FES pricing algorithm with the exploitation rate g defined by Eq. (2), then, for any valuation $v \in [0, 1]$ and $T \ge 4$, the truthful regret is upper bounded:*

$$\mathrm{TReg}(T, \mathcal{A}, v) \le \left(v + \frac{3}{2}\right)(\log_2 \log_2 T + 2). \quad (3)$$

PROOF. Let L be the number of phases conducted by the algorithm during T rounds, then we decompose the total regret over T rounds into the sum of the phases' regrets: $\mathrm{TReg}(T, \mathcal{A}, v) = \sum_{l=0}^{L} R_l$. For the regret at each phase except the last one, the following equality holds

$$R_l = \sum_{k=1}^{K_l}(v - p_{l,k}) + v + g(l)(v - p_{l,K_l}), \quad l = 0, \dots, L-1,$$

where the first, second, and third terms correspond to the exploration rounds with acceptance, the reject round, and the exploitation rounds, respectively. Since the price p_{l,K_l+1} is rejected, then we have $v < p_{l,K_l+1}$ (the buyer is truthful), $v \in [p_{l,K_l}, p_{l,K_l} + \epsilon_l)$, and $p_{l+1,k} \in [p_{l,K_l}, p_{l,K_l} + \epsilon_l) \, \forall k \le K_{l+1} < N_{l+1}$. Hence, for $l = 1, \dots, L$, we have

$$v - p_{l,K_l} < \epsilon_l; \qquad v - p_{l,k} < \epsilon_l(N_l - k) \, \forall k \in \mathbb{Z}_{N_l};$$

$$\text{and} \qquad \sum_{k=1}^{K_l}(v - p_{l,k}) < \epsilon_l \sum_{k=1}^{N_l - 1}(N_l - k) = \frac{1 - \epsilon_{l-1}}{2}.$$

For $l = 0$, one has $\sum_{k=1}^{K_0}(v - p_{0,k}) \le 1/2$. Hence, by Eq. (2),

$$R_l \le \frac{1}{2} + v + g(l) \cdot \epsilon_l \le v + \frac{3}{2}, \quad l = 0, \dots, L-1.$$

Moreover, this inequality holds for the L-th phase, since it differs from the other ones only in possible absence of some rounds (exploration or exploitation ones), but this absence can be easily upper-bounded by the regret of a possible L-th phase as if all these rounds are played in. Finally, one has

$$\mathrm{TReg}(T, \mathcal{A}, v) = \sum_{l=0}^{L} R_l \le \left(v + \frac{3}{2}\right)(L+1).$$

Thus, one needs only to estimate the number of phases L by the number of rounds T. So, $T = \sum_{l=0}^{L-1}(K_l + 1 + g(l)) + K_L + 1 + g_L(L) \ge g(L-1)$, for $T \ge 1 + 1 + g(0)$ (when $v < 1$, otherwise Eq. (3) holds). Hence $g(L-1) = 2^{2^{L-1}} \le T$, which is equivalent to $L \le \log_2 \log_2 T + 1$, and we get Eq. (3). □

5.2 Strategic setting

In this subsection, for the strategic setting, we show, first, that there does not exist a no-regret horizon-independent algorithm in the class $\mathbf{RWC}$ (Theorem 4). Second, we present our no-regret horizon-independent PRRFES algorithm from the class $\mathbf{C_R}$ and prove its optimality (Theorem 5).

The key drawback of a consistent algorithm against a strategic buyer is that he can lie once and due to consistency receive prices at least on ε lower than his valuation v. We formalize that intuition in the following general statement.

THEOREM 4. *For any horizon-independent regular weakly consistent pricing algorithm $\mathcal{A}$ and any $\gamma \in (0, 1)$, there exists a valuation $v \in [0, 1]$ s.t. $\mathrm{SReg}(T, \mathcal{A}, v, \gamma) = \Omega(T)$.*

PROOF SKETCH. If the algorithm $\mathcal{A}$ is not dense, then the theorem holds due to $\mathcal{A} \in \mathbf{RWC}$ and Corollary 1. For a dense algorithm, let us consider the root node $\mathfrak{n}_1 \in \mathfrak{T}(\mathcal{A})$ and the first offered price $p^{\mathfrak{n}_1}$. If $0 < p^{\mathfrak{n}_1} < 1$, we decompose the set of all buyer strategies into three sets $B_0 \sqcup B_- \sqcup B_+$:

- B_0 contains strategies whose price sequences $\{p_t\}_{t=1}^{\infty}$ are constant: $p_t = p^{\mathfrak{n}_1} \, \forall t \in \mathbb{N}$;
- for a strategy from B_-, the price sequence $\{p_t\}_{t=1}^{\infty}$ has the form: $\exists t_0 \in \mathbb{N}$ s.t. $p_{t_0+1} < p_{t_0}$ and $p_t = p^{\mathfrak{n}_1}, t = 1, .., t_0$;
- for a strategy from B_+, its price sequence $\{p_t\}_{t=1}^{\infty}$ has the form: $\exists t_0 \in \mathbb{N}$ s.t. $p_{t_0+1} > p_{t_0}$ and $p_t = p^{\mathfrak{n}_1}, t = 1, .., t_0$.

First, note that $B_- \neq \varnothing$ since, otherwise, the algorithm will be non-dense (due to $p \ge p^{\mathfrak{n}_1} > 0 \forall p \in \wp(\mathcal{A})$). Moreover, since $\mathcal{A}$ is regular weakly consistent, there exists[8] a strategy

[8] To show the existence of $\hat{\mathbf{a}}$, just assume the contrary and use $\mathcal{A} \in \mathbf{RWC}$ to obtain the contradiction with density of $\mathcal{A}$ (it is fairly technical and is missed due to space constraints).

$\hat{\mathbf{a}} \in B_-$ with its price sequence $\{\hat{p}_t\}_{t=1}^{\infty}$ such that

$$\exists t_1 \in \mathbb{N} : \hat{p}_{t_1+1} < \hat{p}_{t_1} < p^{\mathfrak{n}_1} \quad \text{and} \quad a_t = 1\, \forall t > t_1. \qquad (4)$$

Let us denote $\Delta = p^{\mathfrak{n}_1} - \hat{p}_{t_1} > 0$, then, $\forall t \geq t_1$, $\hat{p}_t \leq \hat{p}_{t_1} = p^{\mathfrak{n}_1} - \Delta$ (due to the weak consistency). Hence, on the one hand, the surplus of this strategy followed by a buyer with the valuation $v_\varepsilon := p^{\mathfrak{n}_1} + \varepsilon$ can be lower bounded in the following way, for $T > t_1$:

$$\mathrm{Sur}_\gamma(T, \mathcal{A}, v_\varepsilon, \hat{\mathbf{a}}) \geq \sum_{t=t_1+1}^{T} \gamma^{t-1}(\Delta + \varepsilon) = (\Delta + \varepsilon)\frac{\gamma^{t_1} - \gamma^{T}}{1-\gamma}. \qquad (5)$$

On the other hand, one can upper bound the surplus of a strategy $\mathbf{a} \in B_+$ followed by a buyer with the valuation v_ε, for $T > 0$, since $p_t \geq p^{\mathfrak{n}_1}\ \forall t \in \mathbb{N}$:

$$\mathrm{Sur}_\gamma(T, \mathcal{A}, v_\varepsilon, \mathbf{a}) \leq \sum_{t=1}^{T} \gamma^{t-1}\varepsilon = \varepsilon\frac{1-\gamma^T}{1-\gamma} \quad \forall \mathbf{a} \in B_+.$$

Let $\varepsilon_0 := \min\{\Delta\gamma^{t_1}(1-\gamma)/(1-\gamma^{t_1}), 1 - p^{\mathfrak{n}_1}\}$, then, $\forall \varepsilon \in (0, \varepsilon_0)$, first, $v_{\varepsilon_0} \in (0,1)$, second, $\varepsilon \leq \Delta(\gamma^{t_1} - \gamma^T)/(1 - \gamma^{t_1})\ \forall T > t_1$, and, hence, the right-hand side of Eq. (5) is larger than the one of Eq. (5), i.e., $\mathrm{Sur}_\gamma(T, \mathcal{A}, v_\varepsilon, \mathbf{a}) < \mathrm{Sur}_\gamma(T, \mathcal{A}, v_\varepsilon, \hat{\mathbf{a}})\ \forall \mathbf{a} \in B_+$.

Thus, we showed that, for $T > t_1$, there exists a strategy in B_- (namely, $\hat{\mathbf{a}}$) that is better (in terms of discounted surplus) than any strategy in B_+ for the buyer with the valuation $v_\varepsilon = p^{n_1} + \varepsilon, \varepsilon \in (0, \varepsilon_0)$. Therefore, the optimal strategy $\mathbf{a}^{\mathrm{Opt}}$ must belong to either B_0 or B_- for $T > t_1$. But, for any strategy $\mathbf{a}$ from $B_0 \cup B_-$, one can lower bound the regret by

$$\mathrm{Reg}(T, \mathcal{A}, v_\varepsilon, \mathbf{a}) \geq \sum_{t:a_t=0} v_\varepsilon + \sum_{t:a_t=1} (v_\varepsilon - p^{\mathfrak{n}_1}) \geq T\varepsilon,$$

and, hence, the strategic regret: $\mathrm{SReg}(T, \mathcal{A}, v_\varepsilon, \gamma) \geq T\varepsilon$ for $T > t_1$. This lower bound is $\Omega(T)$ since ε and t_1 are independent of T. Finally, the case of $p^{\mathfrak{n}_1} = 0$ or 1 can be reduced to the previously considered case (through replacing the first node $\mathfrak{n}_1$ by some node $\tilde{\mathfrak{n}} \in \mathfrak{T}(\mathcal{A})$ s.t. $p^{\tilde{\mathfrak{n}}} \in (0,1)$), which is fairly technical and is missed due to space constraints. □

Remark 1. Theorem 4 holds for some weakly consistent algorithms other than only regular ones (regularity is only used when we prove the existence of $\hat{\mathbf{a}}$ satisfying Eq.(4) and make the reduction of the cases $p^{\mathfrak{n}_1} = 0, 1$). However, the research question on the existence of a no-regret horizon-independent algorithm in the class **WC** remains open.

Before presenting our best algorithm whose strategic regret is $O(\log\log T)$, note that the technique of penalization rounds introduced in the algorithm PFS (see Sec. 3.3 and [37]) cannot alone improve a horizon-independent consistent algorithm to a no-regret pricing due to Theorem 4, since the modification will belong to the class **RWC**, and any attempt with straightforward injections of penalization rounds to our algorithm FES will thus be unsuccessful.

So, we go to seek for a desirable algorithm beyond this class **RWC** (which is poor to contain a no-regret algorithm in the strategic setting) and relax the left consistency assumption by considering the class $\mathbf{C_R}$ (see Def. 6). We remain the right-side assumption since the optimal buyer strategy is still non-losing one, i.e., the buyer never lies when he accepts a price (see the discussion after Lemma 1).

Algorithm 2 Pseudo-code of the PRRFES algorithm

```
 1: Input: $r \in \mathbb{N}$ and $g : \mathbb{Z}_+ \to \mathbb{Z}_+$
 2: Initialize: $q := 0$, $p := 1/2$, $l := 0$
 3: while the buyer plays do
 4:   Offer the price $p$ to the buyer
 5:   if the buyer accepts the price then
 6:     $q := p$
 7:   else
 8:     Offer the price $p$ to the buyer for $r - 1$ rounds
 9:     if the buyer accepts one of the prices then
10:       go to line 6
11:     end if
12:     Offer the price $q$ to the buyer for $g(l)$ rounds
13:     $l := l + 1$
14:   end if
15:   if $p < 1$ then
16:     $p := q + 2^{-2^l}$
17:   end if
18: end while
```

Let $\delta^l_{\mathfrak{n}} := p^{\mathfrak{n}} - \inf_{\mathfrak{m} \in \mathfrak{L}(\mathfrak{n})} p^{\mathfrak{m}}$ be the left increment [37], then the following proposition (which is an analogue of the one from [37] obtained for the fully consistent case) holds

Proposition 2. *Let $\gamma \in (0,1)$, $\mathcal{A}$ be a pricing algorithm, $\mathfrak{n} \in \mathfrak{T}(\mathcal{A})$ be a starting node in a r-length penalization sequence (see Def. 2), and $r > \log_\gamma(1-\gamma)$. If the price $p^{\mathfrak{n}}$ is rejected by the strategic buyer, then the following inequality on his valuation v holds:*

$$v - p^{\mathfrak{n}} < \zeta_{r,\gamma}\delta^l_{\mathfrak{n}}, \quad \textit{where} \quad \zeta_{r,\gamma} := \frac{\gamma^r}{1 - \gamma - \gamma^r}. \qquad (6)$$

Proof. For each node $\mathfrak{m} \in \mathfrak{T}(\mathcal{A})$, let $S(\mathfrak{m})$ be the surplus obtained by the buyer when playing an optimal strategy against $\mathcal{A}$ after reaching the node $\mathfrak{m}$. Since the price $p^{\mathfrak{n}}$ is rejected then the following inequality holds [37, Lemma 1]

$$\gamma^{t^{\mathfrak{n}}-1}(v - p^{\mathfrak{n}}) + S(\mathfrak{r}(\mathfrak{n})) < S(\mathfrak{l}(\mathfrak{n})). \qquad (7)$$

The surplus $S(\mathfrak{r}(\mathfrak{n}))$ is lower bounded by 0, while the left subtree's surplus $S(\mathfrak{l}(\mathfrak{n}))$ can be upper bounded as follows (using $p^{\mathfrak{n}} - p^{\mathfrak{m}} \leq \delta^l_{\mathfrak{n}}\ \forall \mathfrak{m} \in \mathfrak{L}(\mathfrak{n})$):

$$S(\mathfrak{l}(\mathfrak{n})) \leq \sum_{t=t^{\mathfrak{n}}+r}^{T} \gamma^{t-1}(v - p^{\mathfrak{n}} + \delta^l_{\mathfrak{n}}) < \frac{\gamma^{t^{\mathfrak{n}}+r-1}}{1-\gamma}(v - p^{\mathfrak{n}} + \delta^l_{\mathfrak{n}}),$$

We plug these bounds in Eq. (7), divide by $\gamma^{t^{\mathfrak{n}}-1}$, and obtain

$$(v - p^{\mathfrak{n}})\left(1 - \frac{\gamma^r}{1-\gamma}\right) < \frac{\gamma^r}{1-\gamma}\delta^l_{\mathfrak{n}},$$

that implies Eq. (6), since $r > \log_\gamma(1-\gamma)$. □

For a right-consistent algorithm $\mathcal{A}$, the increment $\delta^l_{\mathfrak{n}}$ is bounded by the difference between the current node's price $p^{\mathfrak{n}}$ and the last accepted price q before reaching this node. Hence, the inequality Eq. (6) give us an insight on *how to guarantee no-lies* for a certain v on a particular round: the closer an offered price is to the last accepted price the smaller the interval of possible valuations v, holding which the strategic buyer may lie on this offer, $v - p^{\mathfrak{n}} < \zeta_{r,\gamma}(p^{\mathfrak{n}} - q)$.

PRRFES algorithm. We improve our algorithm FES designed for truthful setting to avoid the causes of a linear regret showed in Theorem 4 and to make him thus robust against a strategic buyer: additionally to options (a) and (b) of FES, we (c) use penalization rounds after a rejection, forcing thus the buyer to lie less (similarly to [37]), and (d) regularly revise rejected prices. Namely, *the Penalized Reject-Revising Fast Exploiting Search pricing algorithm* (*PRRFES*) works in phases initialized by the phase index $l := 0$, the last accepted price before the current phase $q_0 := 0$, the iteration parameter $\epsilon_0 := 1/2$, and the number of offers $N_0 := 2$; at each phase $l \in \mathbb{Z}_+$, it sequentially offers prices $p_{l,k} := q_l + k\epsilon_l, k \in \mathbb{N}$ (i.e., in contrast to FES, k can now be higher than N_l, thus, it can explore prices higher than the earlier rejected one $p_{l,N_l} = p_{l-1,K_{l-1}+1}$), with ϵ_l and N_l defined in Eq. (1); if a price $p_{l,k}$ with $k = K_l + 1 \geq 1$ is rejected, (1) it offers this price p_{l,K_l+1} for $r - 1$ rounds (penalization: if one of them is accepted, PRRFES continues offering $p_{l,k}, k = K_l + 2, ..$ following the Definition 2), (2) it offers the price p_{l,K_l} for $g(l)$ rounds (exploitation), and (3) PRRFES goes to the next phase by setting $q_{l+1} := p_{l,K_l}$ and $l := l + 1$. The pseudo-code of PRRFES is presented in Alg. 2, which is in the class $\mathbf{C_R}$.

Theorem 5. *Let $\gamma_0 \in (0, 1)$ and $\mathcal{A}$ be the PRRFES pricing algorithm with $r \geq r_{\gamma_0} := \lceil \log_{\gamma_0}((1-\gamma_0)/2) \rceil$ and the exploitation rate g defined by Eq. (2), then, for any valuation $v \in [0, 1]$ and $T \geq 4$, the strategic regret is upper bounded:*

$$\mathrm{SReg}(T, \mathcal{A}, v, \gamma) \leq (rv+4)(\log_2 \log_2 T+2) \;\forall\gamma \in (0, \gamma_0]. \quad (8)$$

Proof. The proof is fairly similar to the one of Theorem 3: the *key difference* is that we exploit:

1. the inequality $v < p_{l,K_l+1} + \epsilon_l$ (which follows from Prop. 2) instead of $v < p_{l,K_l+1}$ of the truthful setting;
2. by the former inequality, the number of accepted prices K_l at each phase l is limited by $2N_l$ instead of $K_l < N_l$ for the truthful setting.

So, decompose the regret $\mathrm{SReg}(T, \mathcal{A}, v, \gamma) = \sum_{l=0}^{L} R_l$, where L is the number of phases during T rounds. For the regret R_l at each phase except the last one, we have

$$R_l = \sum_{k=1}^{K_l}(v - p_{l,k}) + rv + g(l)(v - p_{l,K_l}), \quad l = 0, \ldots, L-1,$$

where the first, second, and third terms correspond to the exploration rounds with acceptance, the reject-penalization rounds, and the exploitation rounds, respectively. First, since the price p_{l,K_l} is 0 or has been accepted, we have $p_{l,K_l} \leq v$ (the optimal strategy is non-losing one for $\mathcal{A} \in \mathbf{C_R}$). Second, since the price p_{l,K_l+1} is rejected, we have $v - p_{l,K_l+1} < p_{l,K_l+1} - p_{l,K_l} = \epsilon_l$ (by Proposition 2 since $\zeta_{r,\gamma_0} < 1$ for $r \geq r_{\gamma_0}$). Hence, the valuation $v \in [p_{l,K_l}, p_{l,K_l} + 2\epsilon_l)$ and all accepted prices $p_{l+1,k}$, $\forall k \leq K_{l+1}$ from the next phase $l + 1$ satisfy:

$$p_{l+1,k} \in [q_{l+1}, v) \subseteq [p_{l,K_l}, p_{l,K_l} + 2\epsilon_l) \quad \forall k \leq K_{l+1},$$

inferring $K_{l+1} < 2N_{l+1}$. For $l = 1, \ldots, L$,

$$v - p_{l,K_l} < 2\epsilon_l; \quad v - p_{l,k} < \epsilon_l(2N_l - k) \,\forall k \in \mathbb{Z}_{2N_l};$$

$$\text{and} \qquad \sum_{k=1}^{K_l}(v - p_{l,k}) < \epsilon_l \sum_{k=1}^{2N_l-1}(2N_l - k) = 2 - \epsilon_{l-1}.$$

For $l = 0$, one has $\sum_{k=1}^{K_0}(v - p_{0,k}) \leq 1/2$. Hence, by Eq. (2),

$$R_l \leq 2 + rv + g(l) \cdot 2\epsilon_l \leq rv + 4, \quad l = 0, \ldots, L-1,$$

and, similarly to the proof of Theorem 3, we get Eq. (8). □

Table 1: Summary on best known regret bounds for different classes of horizon-independent algorithms (the ones in blue are contributed by our study).

Scenario\Alg. class	**SC**	**C**	**RWC**	**WC**	**RC**	Any
Truthful	$\Omega(T)$	$\Theta(\log\log T)$ by FES				
Strategic, $\gamma \in (0, 1)$	$\Omega(T)$			open quest.	$\Theta(\log\log T)$ by PRRFES	
Strategic, $\gamma = 1$	$\Omega(T)$					

5.3 Discussion and summary

One algorithm for both scenarios. It is easy to see that the upper bound in Eq. (8) holds for the truthful regret TReg of PRRFES. Therefore, this algorithm can be applied against both truthful (myopic) and strategic buyers without a priori knowledge of what type of buyer the seller is facing.

Strategic buyer with infinite horizon. Note that the proofs of Proposition 2 and, thus, Theorem 5 do not exploit the finiteness of the buyer's horizon. Hence, the upper bound in Eq. (8) holds for the case, when the buyer selects the optimal strategy $\mathbf{a}^{\mathrm{Opt}}$ so as to maximize his surplus over infinite number of rounds, i.e., $\mathrm{Sur}_\gamma(\infty, \mathcal{A}, v, \mathbf{a})$, (being motivated by the fact that the seller can play infinitely due to utilization of a horizon-independent algorithm) and PRRFES can be applied against strategic buyers with infinite horizon.

Summary on regret bounds. In Table 1, we summarize all best known regret bounds for different classes of horizon-independent algorithms. In each cell, we indicate either a tight regret bound with the algorithm by which the bound is achieved from the corresponding class, or a linear lower bound if there does not exist a no-regret algorithm from the corresponding class. We remind that the research question on the existence of a no-regret horizon-independent algorithm in the class **WC** remains open.

6. CONCLUSIONS

We studied horizon-independent online learning algorithms in the scenario of repeated posted-price auctions with a strategic buyer that holds a fixed private valuation. First, we closed the gap between the previously best known upper and lower bounds on strategic regret. Second, we presented the novel horizon-independent algorithm that can be applied both against strategic and truthful buyers with a tight regret bound in $\Theta(\log\log T)$, outperforming the previously known algorithms (even in the horizon-independent variants obtained by a state-of-the-art technique). Finally, we provided a thorough theoretical analysis of several broad families of pricing algorithms, that may help in future studies on a more sophisticated scenarios and auction mechanisms.

7. REFERENCES

[1] D. Agarwal, S. Ghosh, K. Wei, and S. You. Budget pacing for targeted online advertisements at linkedin. In *KDD'2014*, pages 1613–1619, 2014.

[2] G. Aggarwal, G. Goel, and A. Mehta. Efficiency of (revenue-) optimal mechanisms. In *EC'2009*, pages 235–242, 2009.

[3] G. Aggarwal, S. Muthukrishnan, D. Pál, and M. Pál. General auction mechanism for search advertising. In *WWW'2009*, pages 241–250, 2009.

[4] K. Amin, M. Kearns, and U. Syed. Bandits, query learning, and the haystack dimension. In *COLT*, pages 87–106, 2011.

[5] K. Amin, A. Rostamizadeh, and U. Syed. Learning prices for repeated auctions with strategic buyers. In *NIPS'2013*, pages 1169–1177, 2013.

[6] K. Amin, A. Rostamizadeh, and U. Syed. Repeated contextual auctions with strategic buyers. In *NIPS'2014*, pages 622–630, 2014.

[7] I. Ashlagi, C. Daskalakis, and N. Haghpanah. Sequential mechanisms with ex-post participation guarantees. In *EC'2016*, 2016.

[8] I. Ashlagi, B. G. Edelman, and H. S. Lee. Competing ad auctions. *Harvard Business School NOM Unit Working Paper*, (10-055), 2013.

[9] M. Babaioff, S. Dughmi, R. Kleinberg, and A. Slivkins. Dynamic pricing with limited supply. *ACM Transactions on Economics and Computation*, 3(1):4, 2015.

[10] Y. Bachrach, S. Ceppi, I. A. Kash, P. Key, and D. Kurokawa. Optimising trade-offs among stakeholders in ad auctions. In *EC'2014*, pages 75–92, 2014.

[11] S. Balseiro, O. Besbes, and G. Y. Weintraub. Dynamic mechanism design with budget constrained buyers under limited commitment. In *EC'2016*, 2016.

[12] S. R. Balseiro, O. Besbes, and G. Y. Weintraub. Repeated auctions with budgets in ad exchanges: Approximations and design. *Management Science*, 61(4):864–884, 2015.

[13] C. Borgs, J. Chayes, N. Immorlica, K. Jain, O. Etesami, and M. Mahdian. Dynamics of bid optimization in online advertisement auctions. In *WWW'2007*, pages 531–540, 2007.

[14] L. E. Celis, G. Lewis, M. M. Mobius, and H. Nazerzadeh. Buy-it-now or take-a-chance: a simple sequential screening mechanism. In *WWW'2011*, pages 147–156, 2011.

[15] N. Cesa-Bianchi, C. Gentile, and Y. Mansour. Regret minimization for reserve prices in second-price auctions. In *SODA'2013*, pages 1190–1204, 2013.

[16] D. Charles, N. R. Devanur, and B. Sivan. Multi-score position auctions. In *WSDM'2016*, pages 417–425, 2016.

[17] X. Chen and Z. Wang. Bayesian dynamic learning and pricing with strategic customers. *Available at SSRN 2715730*, 2016.

[18] Y. Chen and V. F. Farias. Robust dynamic pricing with strategic customers. In *EC'2015*, pages 777–777, 2015.

[19] M. Chhabra and S. Das. Learning the demand curve in posted-price digital goods auctions. In *ICAAMS'2011*, pages 63–70, 2011.

[20] M. C. Cohen, I. Lobel, and R. Paes Leme. Feature-based dynamic pricing. In *EC'2016*, 2016.

[21] A. V. den Boer. Dynamic pricing and learning: historical origins, current research, and new directions. *Surveys in operations research and management science*, 20(1):1–18, 2015.

[22] P. Dütting, M. Henzinger, and I. Weber. An expressive mechanism for auctions on the web. In *WWW'2011*, pages 127–136, 2011.

[23] B. Edelman and M. Ostrovsky. Strategic bidder behavior in sponsored search auctions. *Decision support systems*, 43(1):192–198, 2007.

[24] G. Goel and M. R. Khani. Revenue monotone mechanisms for online advertising. In *WWW'2014*, pages 723–734, 2014.

[25] R. Gomes and V. Mirrokni. Optimal revenue-sharing double auctions with applications to ad exchanges. In *WWW'2014*, pages 19–28, 2014.

[26] D. He, W. Chen, L. Wang, and T.-Y. Liu. A game-heoretic machine learning approach for revenue maximization in sponsored search. In *IJCAI'2013*, pages 206–212, 2013.

[27] H. Heidari, M. Mahdian, U. Syed, S. Vassilvitskii, and S. Yazdanbod. Pricing a low-regret seller. In *ICML'2016*, pages 2559–2567, 2016.

[28] P. Hummel and P. McAfee. Machine learning in an auction environment. In *WWW'2014*, pages 7–18, 2014.

[29] K. Iyer, R. Johari, and M. Sundararajan. Mean field equilibria of dynamic auctions with learning. *ACM SIGecom Exchanges*, 10(3):10–14, 2011.

[30] Y. Kanoria and H. Nazerzadeh. Dynamic reserve prices for repeated auctions: Learning from bids. *Available at SSRN 2444495*, 2014.

[31] R. Kleinberg and T. Leighton. The value of knowing a demand curve: Bounds on regret for online posted-price auctions. In *Foundations of Computer Science*, pages 594–605, 2003.

[32] V. Krishna. *Auction theory*. Academic press, 2009.

[33] T. Lin, J. Li, and W. Chen. Stochastic online greedy learning with semi-bandit feedbacks. In *NIPS'2015*, pages 352–360, 2015.

[34] B. Lucier, R. Paes Leme, and E. Tardos. On revenue in the generalized second price auction. In *WWW'2012*, pages 361–370, 2012.

[35] M. Mohri and A. M. Medina. Learning theory and algorithms for revenue optimization in second price auctions with reserve. In *ICML'2014*, pages 262–270, 2014.

[36] M. Mohri and A. M. Medina. Non-parametric revenue optimization for generalized second price auctions. In *UAI'2015*, 2015.

[37] M. Mohri and A. Munoz. Optimal regret minimization in posted-price auctions with strategic buyers. In *NIPS'2014*, pages 1871–1879, 2014.

[38] M. Mohri and A. Munoz. Revenue optimization against strategic buyers. In *NIPS'2015*, pages 2530–2538, 2015.

[39] J. H. Morgenstern and T. Roughgarden. On the pseudo-dimension of nearly optimal auctions. In *NIPS'2014*, pages 136–144, 2015.

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

[41] G. Noti, N. Nisan, and I. Yaniv. An experimental evaluation of bidders' behavior in ad auctions. In *WWW'2014*, pages 619–630, 2014.

[42] M. Ostrovsky and M. Schwarz. Reserve prices in internet advertising auctions: A field experiment. In *EC'2011*, pages 59–60, 2011.

[43] R. Paes Leme, M. Pál, and S. Vassilvitskii. A field guide to personalized reserve prices. In *WWW'2016*, pages 1093–1102, 2016.

[44] T. Roughgarden and J. R. Wang. Minimizing regret with multiple reserves. In *EC'2016*, pages 601–616, 2016.

[45] M. R. Rudolph, J. G. Ellis, and D. M. Blei. Objective variables for probabilistic revenue maximization in second-price auctions with reserve. In *WWW'2016*, pages 1113–1122, 2016.

[46] Y. Sun, Y. Zhou, and X. Deng. Optimal reserve prices in weighted gsp auctions. *Electronic Commerce Research and Applications*, 13(3):178–187, 2014.

[47] D. R. Thompson and K. Leyton-Brown. Revenue optimization in the generalized second-price auction. In *EC'2013*, pages 837–852, 2013.

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

[49] H. R. Varian. Online ad auctions. *The American Economic Review*, 99(2):430–434, 2009.

[50] H. R. Varian and C. Harris. The vcg auction in theory and practice. *The Amer. Econ. Rev.*, 104(5):442–445, 2014.

[51] J. Weed, V. Perchet, and P. Rigollet. Online learning in repeated auctions. *JMLR*, 49:1–31, 2016.

[52] S. Yuan, J. Wang, B. Chen, P. Mason, and S. Seljan. An empirical study of reserve price optimisation in real-time bidding. In *KDD'2014*, pages 1897–1906, 2014.

[53] Y. Zhu, G. Wang, J. Yang, D. Wang, J. Yan, J. Hu, and Z. Chen. Optimizing search engine revenue in sponsored search. In *SIGIR'2009*, pages 588–595, 2009.

[54] M. Zoghi, Z. S. Karnin, S. Whiteson, and M. De Rijke. Copeland dueling bandits. In *NIPS'2015*, pages 307–315, 2015.