Weakly Consistent Optimal Pricing Algorithms in Repeated Posted-Price Auctions with Strategic Buyer

Alexey Drutsa [1 2]

Abstract

We study revenue optimization learning algorithms for repeated posted-price auctions where a seller interacts with a single strategic buyer that holds a fixed private valuation for a good and seeks to maximize his cumulative discounted surplus. We propose a novel algorithm that never decreases offered prices and has a tight strategic regret bound of $\Theta(\log\log T)$. This result closes the open research question on the existence of a no-regret horizon-independent weakly consistent pricing. We also show that the property of non-decreasing prices is nearly necessary for a weakly consistent algorithm to be a no-regret one.

1. Introduction

Real-time ad exchanges, search engines, social networks, and other Internet companies consider *revenue maximization* as one of the most important directions for development of their online advertising platforms (Gomes & Mirrokni, 2014; Balseiro et al., 2015; Charles et al., 2016; Agarwal et al., 2014). A large part of advertisement inventory is sold via widely applicable second price auctions (He et al., 2013; Mohri & Medina, 2014), including their generalizations such as GSP (Varian, 2007; Sun et al., 2014) and Vickrey-Clarke-Groves (VCG) (Varian, 2009; Varian & Harris, 2014) auctions. The optimization of revenue here is mostly controlled by means of reserve prices, whose proper setting is studied both by game-theoretical methods (Myerson, 1981; Krishna, 2009) and by machine learning approaches (Nisan et al., 2007; Cesa-Bianchi et al., 2013; Paes Leme et al., 2016). A large number of online auctions run, for instance, by ad exchanges involve only a single advertiser (Amin et al., 2013; Mohri & Munoz, 2014; Drutsa, 2017), and, in this case, a second-price auction with reserve is equivalent to a *posted-price auction* (Kleinberg & Leighton, 2003) where the seller sets a reserve price for a good (for instance, an ad space) and the buyer decides whether to accept or reject this price i.e., to bid above or below it).

In this work, we focus on a scenario when the seller *repeatedly* interacts through a posted-price mechanism with the *same* strategic buyer that holds a *fixed* private valuation for a good and seeks to maximize his cumulative discounted surplus (Amin et al., 2013). At each round of this game, the seller is able to choose the price based on previous decisions of the buyer: i.e., she applies a deterministic online learning algorithm, which is announced to the buyer in advance. The seller's goal is to maximize her cumulative revenue over a finite time horizon T, which is generally reduced to *regret* minimization[1]. Thus, the seller seeks for a *no-regret* pricing algorithm, i.e., the one with a sublinear regret on T (Mohri & Munoz, 2014; Amin et al., 2014; Chen & Wang, 2016).

For this setting, the algorithm PRRFES with tight strategic regret bound of $\Theta(\log\log T)$ was recently proposed (Drutsa, 2017). This algorithm is horizon-independent and right-consistent (i.e., it never proposes prices lower than earlier accepted ones). However, its key drawback is its ability to decrease an offered price after its rejection, but then to propose higher prices than rejected one in subsequent rounds (not satisfying thus the left consistency). Such a behavior of the algorithm may be confusing for a buyer: he may doubt that the announced algorithm is used by the seller[2]. The full consistency (both right, and left) is thus quite welcome. Formally, there does not exist a no-regret horizon-independent consistent algorithm (Drutsa, 2017). But, it is still an open question whether there exists a *weakly* consistent algorithm with these properties. Weak consistency is a slightly relaxed variant of consistency and means, in particular, that, once the seller decreases a price after a rejection, the future prices will never exceed the rejected one.

In our study, we propose a novel algorithm that never decreases offered prices and can be applied against strategic buyers with a tight regret bound of $\Theta(\log\log T)$ (Th. 1).

[1]Yandex, Moscow, Russia [2]Faculty of Mechanics and Mathematics, Lomonosov Moscow State University, Moscow, Russia. Correspondence to: Alexey Drutsa <adrutsa@yandex.ru>.

Proceedings of the 35^{th} International Conference on Machine Learning, Stockholm, Sweden, PMLR 80, 2018.

[1]We study the *worst-case* regret minimization following (Kleinberg & Leighton, 2003; Mohri & Munoz, 2014; Drutsa, 2017).

[2]See Appendix C in Supp.Mat. for an example from practice.

This result constitutes *the main contribution of our work* and closes the open research question on the existence of a no-regret horizon-independent pricing in the class of weakly consistent algorithms. We also show that the property of non-decreasing prices is crucial and, in fact, is nearly necessary for a weakly consistent algorithm to be a no-regret one: namely, a double decrease of offered prices by a weakly consistent algorithm is enough to cause a linear regret (Th. 2). Construction and analysis of the proposed algorithm have required introduction of novel techniques, which are contributed by our work as well. They include a novel transformation which maps a right-consistent algorithm to a both weakly and right- consistent one that is only able to increase prices (Sec. 3); and methods that guarantee acceptance of exploitation prices that have not been earlier accepted (Sec. 4).

2. Preliminaries

2.1. Setup of Repeated Posted-Price Auctions

We study the mechanism of *repeated posted-price auctions* earlier considered, e.g., in (Amin et al., 2013; Mohri & Munoz, 2014; Drutsa, 2017). Namely, the seller repeatedly proposes goods (e.g., advertisement opportunities) to a single buyer over T rounds (the time horizon): one good per round. The buyer holds a *fixed private valuation* $v \in [0;1]$ for a good, i.e., the valuation v is unknown to the seller and is equal for goods offered in all rounds. At each round $t \in \{1, \ldots, T\}$, 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 $t \in \{1, \ldots, T\}$ can depend only on the buyer's actions during the previous rounds $\mathbf{a}_{1:t-1}$[3]. Following (Drutsa, 2017), we are studying algorithms that do not depend on the horizon T since it is very natural in practice (e.g., of ad exchanges) that the seller does not know in advance the number of rounds T that the buyer wants to interact with him. Let $\mathbf{A}$ be the set of such algorithms.

Hence, given a pricing algorithm $\mathcal{A} \in \mathbf{A}$, a buyer strategy $\mathbf{a} = \{a_t\}_{t=1}^T$ uniquely defines the corresponding price sequence $\{p_t\}_{t=1}^T$, which, in turn, determines the seller's total revenue $\sum_{t=1}^T a_t p_t$. 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 (Kleinberg & Leighton, 2003; Amin et al., 2013; Mohri & Munoz, 2014; Drutsa, 2017). 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 $\mathrm{Reg}(T, \mathcal{A}, v, \mathbf{a}) := \sum_{t=1}^T (v - a_t p_t)$.

Following a standard assumption in mechanism design that matches the practice in ad exchanges (Mohri & Munoz, 2014), the seller's pricing algorithm $\mathcal{A}$ *is announced to the buyer in advance*. The buyer can then act strategically against this algorithm and is assumed to follow the optimal strategy $\mathbf{a}^{\mathrm{Opt}}(T, \mathcal{A}, v, \boldsymbol{\gamma})$ that maximizes the buyer's $\boldsymbol{\gamma}$-discounted surplus (Amin et al., 2013):

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

i.e., $\mathbf{a}^{\mathrm{Opt}}(T, \mathcal{A}, v, \boldsymbol{\gamma}) := \mathrm{argmax}_{\mathbf{a}} \mathrm{Sur}_{\boldsymbol{\gamma}}(T, \mathcal{A}, v, \mathbf{a})$, where $\boldsymbol{\gamma} = \{\gamma_t\}_{t=1}^\infty$ is *the discount sequence*, which is positive ($\gamma_t > 0 \; \forall t \in \mathbb{N}$) and has convergent sums ($\sum_{t=1}^\infty \gamma_t < \infty$). 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

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

Thus, we consider a two-player non-zero sum repeated game with commitment, incomplete information, and unlimited supply, which is introduced by Amin et al. (2013) and considered in (Mohri & Munoz, 2014; Drutsa, 2017): the buyer seeks to maximize his surplus, while the seller's objective is to minimize her strategic regret (i.e., maximize her revenue). Note that only the buyer's objective is discounted over time (not the seller's one), which is motivated by the observation that sellers are far more willing to wait for revenue than buyers are willing to wait for goods in important real-world markets like online advertising (Amin et al., 2013; Mohri & Munoz, 2014).

In our setting, following (Kleinberg & Leighton, 2003; Amin et al., 2013; 2014; Mohri & Munoz, 2014; 2015; Drutsa, 2017), we seek for algorithms that attain $o(T)$ strategic regret (i.e., the averaged regret converges to zero when $T \to \infty$) for the worst-case valuation $v \in [0, 1]$. An algorithm $\mathcal{A}$ is said to be a *no-regret* one when $\sup_{v \in [0,1]} \mathrm{Reg}(T, \mathcal{A}, v, \mathbf{a}^{\mathrm{Opt}}) = o(T)$. The optimization goal consists in finding of algorithms with the lowest possible strategic regret upper bound of the form $O(f(T))$. Here we treat their optimality in terms of $f(T)$ with the slowest growth as $T \to \infty$. This means that the averaged regret has the best rate of convergence to zero.

2.2. Notations and Auxiliary Definitions

Similarly to (Drutsa, 2017), a deterministic pricing algorithm $\mathcal{A}$ can be associated with an infinite complete binary tree $\mathfrak{T}(\mathcal{A})$ (since we consider horizon-independent algorithms). Each node $\mathfrak{n} \in \mathfrak{T}(\mathcal{A})$[4] is labeled with the price $p^{\mathfrak{n}}$ offered by $\mathcal{A}$. We denote the node's depth + 1 by $t^{\mathfrak{n}}$. 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 the 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] To simplify notations, $\mathfrak{n} \in \mathfrak{T}$ means "$\mathfrak{n}$ is a node of the tree $\mathfrak{T}$".

$\mathfrak{l}(\mathfrak{n})$ ($\mathfrak{r}(\mathfrak{n})$ resp.) are denoted by $\mathfrak{L}(\mathfrak{n})$ ($\mathfrak{R}(\mathfrak{n})$ resp.). The operators $\mathfrak{l}(\cdot)$ and $\mathfrak{r}(\cdot)$ sequentially applied s times to a node $\mathfrak{n}$ are denoted by $\mathfrak{l}^s(\mathfrak{n})$ and $\mathfrak{r}^s(\mathfrak{n})$ respectively, $s \in \mathbb{N}$. The root node of a tree $\mathfrak{T}$ is denoted by $\mathfrak{e}(\mathfrak{T})$.

So, the algorithm's work flow is as follows: it starts at the root $\mathfrak{e}(\mathfrak{T}(\mathcal{A}))$ of the tree $\mathfrak{T}(\mathcal{A})$ by offering the first price $p^{\mathfrak{e}(\mathfrak{T}(\mathcal{A}))}$ 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 the time horizon T. The pseudo-code of this process is presented in Alg. 1. For a node $\mathfrak{n} \in \mathfrak{T}(\mathcal{A})$, $t^{\mathfrak{n}}$ equals to the round at which the price of this node is offered. Each node $\mathfrak{n} \in \mathfrak{T}(\mathcal{A})$ uniquely determines the buyer decisions up to the round $t^{\mathfrak{n}} - 1$[5].

We say that two infinite complete trees $\mathfrak{T}_1$ and $\mathfrak{T}_2$ are *price equivalent* (and write $\mathfrak{T}_1 \cong \mathfrak{T}_2$) if the trees have the same node labeling when we naturally match the nodes between the trees (starting from the roots): i.e., following the same strategy in both trees, the buyer receives the same sequence of prices. 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 $\mathcal{A}$.

2.3. Background on Pricing Algorithms

First of all, we remind several classes (sets) of algorithms that were introduced in (Mohri & Munoz, 2014; Drutsa, 2017) and include the definitions of pricing consistency of different type, which are actively used in our work. After that, we briefly overview pricing algorithms from existing studies (Kleinberg & Leighton, 2003; Amin et al., 2013; Mohri & Munoz, 2014; Drutsa, 2017).

Notion of consistency. Since the buyer holds a fixed valuation, we could expect that a smart online learning algorithm should work as follows: after an acceptance (a rejection), it should set only no lower (no higher, respectively) prices than the offered one. Formally, this leads to the definition:

Definition 1 ((Mohri & Munoz, 2014))**.** An algorithm $\mathcal{A}$ is said to be *consistent* ($\mathcal{A}$ in the class $\mathbf{C}$) if, for any node $\mathfrak{n} \in \mathfrak{T}(\mathcal{A})$, $p^{\mathfrak{m}} \geq p^{\mathfrak{n}}\ \forall \mathfrak{m} \in \mathfrak{R}(\mathfrak{n})$ and $p^{\mathfrak{m}} \leq p^{\mathfrak{n}}\ \forall \mathfrak{m} \in \mathfrak{L}(\mathfrak{n})$.

A consistent algorithm $\mathcal{A}$ is based on a clear idea (Drutsa, 2017): 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 famous example of a consistent algorithm is the binary search.

Definition 2 ((Drutsa, 2017))**.** An algorithm $\mathcal{A}$ is said to be *weakly consistent* ($\mathcal{A}$ in the class $\mathbf{WC}$) if, for any node $\mathfrak{n} \in \mathfrak{T}(\mathcal{A})$, (a) if $p^{\mathfrak{r}(\mathfrak{n})} \neq p^{\mathfrak{n}}$ then $p^{\mathfrak{m}} \geq p^{\mathfrak{n}}\ \ \forall \mathfrak{m} \in \mathfrak{R}(\mathfrak{n})$; and, (b) if $p^{\mathfrak{l}(\mathfrak{n})} \neq p^{\mathfrak{n}}$ then $p^{\mathfrak{m}} \leq p^{\mathfrak{n}}\ \ \forall \mathfrak{m} \in \mathfrak{L}(\mathfrak{n})$.

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 App.D.1 as well.

Definition 3 ((Drutsa, 2017))**.** 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 higher than a rejected one (in contrast to consistent algorithms). We have the following relations of classes: $\mathbf{C} \subset \mathbf{WC}$ and $\mathbf{C} \subset \mathbf{C_R}$.

Background. The consistency represents a quite reasonable property, when the buyer is myopic (truthful, i.e., $a_t = 1 \Leftrightarrow p_t \leq v$), because a reported buyer decision correctly locates v in $[0, 1]$. Kleinberg & Leighton (2003) showed that the regret of any pricing algorithm against a myopic buyer is lower bounded by $\Omega(\log \log T)$ and proposed a horizon-dependent consistent algorithm *FS* (*Fast Search*), that has tight regret bound of $\Theta(\log \log T)$ against such buyers.

A strategic buyer, incited by surplus maximization, may mislead a consistent pricing algorithm (Amin et al., 2014; Mohri & Munoz, 2014). To overcome this, Mohri & Munoz (2014) proposed to inject so-called *penalization rounds* (see Def. 4) after each rejection into the algorithm FS and got, in this way, the algorithm PFS with strategic regret bound of $O(\log T \log \log T)$. This outperforms the algorithm "Monotone" (Amin et al., 2013) with regret bound of $O(T^{1/2})$. Both algorithms are horizon-dependent and are not optimal.

Definition 4 ((Mohri & Munoz, 2014; Drutsa, 2017))**.** Nodes $\mathfrak{n}_1, \ldots, \mathfrak{n}_r \in \mathfrak{T}(\mathcal{A})$ are said to be a (r-length) penalization sequence if $\mathfrak{n}_{i+1} = \mathfrak{l}(\mathfrak{n}_i)$, $p^{\mathfrak{n}_{i+1}} = p^{\mathfrak{n}_i}$, and $\mathfrak{R}(\mathfrak{n}_{i+1}) \cong \mathfrak{R}(\mathfrak{n}_i), i = 1, \ldots, r - 1$.

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, when the discount sequence $\boldsymbol{\gamma}$ is decreasing.

An optimal pricing was found in (Drutsa, 2017), where horizon-independent algorithms were studied and the causes of a linear regret in different classes of consistent algorithms were analyzed step-by-step. First, the algorithm FES (Drutsa, 2017) was proposed as a modification of the FS by injecting exploitation rounds after each rejection to obtain a consistent horizon-independent algorithm against truthful buyer with tight regret bound of

[5] Therefore, 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 $\mathfrak{e}(\mathfrak{T}(\mathcal{A}))$ and goes to a t-depth node. The strategy prices are in turn the ones that are in the nodes lying along this path.

$\Theta(\log\log T)$. Second, this pricing was upgraded to the algorithm PRRFES (Drutsa, 2017) to act against strategic buyers. Namely, it was shown that there is no no-regret pricing in the class $\mathbf{RWC}$ of regular weakly consistent algorithms (a subset of $\mathbf{WC}$, see Def. A.1 in Suppl.Mat.), which comprises, in particular, all consistent horizon-independent algorithms even being modified by penalization rounds. This led to a guess that possibly the left consistency requirement should be relaxed. Inspired by this guess, the optimal right-consistent algorithm PRRFES with tight strategic regret bound of $\Theta(\log\log T)$ was built. However, the research question on the existence of a no-regret horizon-independent algorithm in the class $\mathbf{WC}$ remained open. Our research goal comprises closing of that open research question.

2.4. Related Work

Most of studies on online advertising auctions lies in the field of game theory (Krishna, 2009; Nisan et al., 2007). A large part of them focused on characterizing equilibrium: efficiency (Aggarwal et al., 2009a), mechanism expressiveness (Dütting et al., 2011), competition across auction platforms (Ashlagi et al., 2013), buyer budget (Agarwal et al., 2014), etc. They considered different auctions (Varian, 2009; Aggarwal et al., 2009b; Celis et al., 2011) and conducted experimental analysis (Ostrovsky & Schwarz, 2011; Thompson & Leyton-Brown, 2013; Lahaie et al., 2018).

Studies on revenue maximization were devoted to both the seller revenue solely (Zhu et al., 2009; He et al., 2013) and various trade-offs as well (Bachrach et al., 2014; Goel & Khani, 2014; Morgenstern & Roughgarden, 2015). The optimization problem was generally reduced to a selection of reserve prices for buyers: for instance, in VCG (Myerson, 1981), GSP (Lucier et al., 2012), and others (Gomes & Mirrokni, 2014; Paes Leme et al., 2016). Reserve prices, in such setups, usually depend on distributions of buyer valuations or bids that thus estimated or fitted (He et al., 2013; Sun et al., 2014; Paes Leme et al., 2016). Alternative approaches learned reserve prices directly (Mohri & Medina, 2014; 2015; Rudolph et al., 2016; Medina & Vassilvitskii, 2017). In contrast to these works, we use an online deterministic learning approach to set prices in 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. An extensive survey on this field is presented in (den Boer, 2015). Dynamic pricing was studied: under game-theoretic view (Iyer et al., 2011; Leme et al., 2012; Chen & Farias, 2015; Balseiro et al., 2016; Ashlagi et al., 2016); as bandit problems (Amin et al., 2011; Zoghi et al.; Lin et al., 2015; Weed et al., 2016); feature-based pricing (Amin et al., 2014; Cohen et al., 2016); and from other aspects (Heidari et al., 2016; Yuan et al., 2014; Roughgarden & Wang, 2016; Feldman et al., 2016; Chawla et al., 2016). Vanunts et al. (2018) study the same repeated auction setup, but seek for algorithms that maximize the strategic revenue *expected* over buyer valuations. A series of studies (Schmidt, 1993; Devanur et al., 2015; Immorlica et al., 2017) close to ours considered repeated sales where the seller does not commit for its pricing policy (in contrast to our setting). *That studies showed that the seller earns less in settings without commitment than with it, what motivates the seller to be interested in commitment.* The most relevant part of these works to ours are (Kleinberg & Leighton, 2003; Amin et al., 2013; Mohri & Munoz, 2014; Drutsa, 2017), where our scenario with a fixed private valuation is considered and whose algorithms are discussed in more details in Sec. 2.3. In contrast to the first three studies, we propose and analyze algorithm that have tight strategic regret bound of $\Theta(\log\log T)$, and, unlike in (Drutsa, 2017), this pricing is weakly consistent and never decreases offered prices.

3. Transformation `pre`

Let us introduce a novel transformation which is referred to as `pre` and transforms any pricing algorithm to another one. First, we define this transformation for labeled binary trees.

Definition 5. Given a non-negative real number $q \in \mathbb{R}_+$ and a labeled binary tree $\mathfrak{T}_1$, the transformation $\mathtt{pre} : (q, \mathfrak{T}_1) \mapsto \mathfrak{T}_2$ is such that the labels (i.e., prices) of the tree $\mathfrak{T}_2$ are defined recursively in the following way starting from the root node $\mathfrak{e}(\mathfrak{T}_2)$ of the tree $\mathfrak{T}_2$:

$$p^{\mathfrak{e}(\mathfrak{T}_2)} := q, \qquad \mathfrak{L}\big(\mathfrak{e}(\mathfrak{T}_2)\big) \cong \mathtt{pre}\Big(q, \mathfrak{L}\big(\mathfrak{e}(\mathfrak{T}_1)\big)\Big)$$
$$\text{and} \quad \mathfrak{R}\big(\mathfrak{e}(\mathfrak{T}_2)\big) \cong \mathtt{pre}\Big(p^{\mathfrak{e}(\mathfrak{T}_1)}, \mathfrak{R}\big(\mathfrak{e}(\mathfrak{T}_1)\big)\Big). \tag{1}$$

Second, since each pricing algorithm $\mathcal{A} \in \mathbf{A}$ is associated with a complete binary tree $\mathfrak{T}(\mathcal{A})$, the transformation `pre` is thus correctly defined for pricing algorithms: namely, $\mathtt{pre} : \mathbb{R}_+ \times \mathbf{A} \to \mathbf{A}$ and $\mathtt{pre}(q, \mathcal{A}_1)$ is the pricing algorithm associated with the tree $\mathtt{pre}(q, \mathfrak{T}(\mathcal{A}_1))$. In Algorithm 2, for better understanding, we provide a reader with a pseudocode that applies the pricing $\mathtt{pre}(q, \mathcal{A})$ with given $q \in \mathbb{R}$ and a source pricing $\mathcal{A} \in \mathbf{A}$[6]. See an example in App.D.2.

Informally speaking, this transformation tracks over the nodes in the source algorithm's tree $\mathfrak{T}(\mathcal{A})$, but, being in a current node $\mathfrak{n} \in \mathfrak{T}(\mathcal{A})$, it offers the price from one of preceding nodes, where the buyer purchased a good last time (or q if never purchased), instead of offering the price $p^{\mathfrak{n}}$ from the current node $\mathfrak{n}$. From the buyer's point of view, the choice between the pricing of the subtree $\mathfrak{L}(\mathfrak{n})$ and the one of the subtree $\mathfrak{R}(\mathfrak{n})$, $\mathfrak{n} \in \mathfrak{T}(\mathcal{A})$, should be made at the round *pre*vious to the one where the price $p^{\mathfrak{n}}$ will be offered.

[6] Alg. 2 is placed side-by-side along with Alg. 1 in order to show the difference between the work flow of the transformed pricing $\mathtt{pre}(q, \mathcal{A})$ and the one of the source pricing $\mathcal{A}$.

Algorithm 1 Pseudo-code of an algorithm $\mathcal{A}$.

Input: $\mathcal{A} \in \mathbf{A}$
Initialize: $\mathfrak{n} := \mathfrak{e}(\mathfrak{T}(\mathcal{A}))$,
while the buyer plays **do**
  Offer the price $p^{\mathfrak{n}}$ to the buyer
  if the buyer accepts the price **then**
    $\mathfrak{n} := \mathfrak{r}(\mathfrak{n})$
  else
    $\mathfrak{n} := \mathfrak{l}(\mathfrak{n})$
  end if
end while

Algorithm 2 Pseudo-code of its transformation $\mathtt{pre}(q, \mathcal{A})$.

Input: $q \in \mathbb{R}$ and $\mathcal{A} \in \mathbf{A}$
Initialize: $\mathfrak{n} := \mathfrak{e}(\mathfrak{T}(\mathcal{A}))$, $p := q$
while the buyer plays **do**
  Offer the price p to the buyer
  if the buyer accepts the price **then**
    $p := p^{\mathfrak{n}}$
    $\mathfrak{n} := \mathfrak{r}(\mathfrak{n})$
  else
    $\mathfrak{n} := \mathfrak{l}(\mathfrak{n})$
  end if
end while

The following key properties hold for this transformation.

Lemma 1. *Let $\mathcal{A} \in \mathbf{C_R}$ be a right-consistent pricing algorithm and $q = \inf \wp(\mathcal{A})$ be the infimum of the algorithm prices, then the transformed pricing algorithm $\mathtt{pre}(q, \mathcal{A})$ is both right-consistent and weakly consistent, i.e., $\mathtt{pre}(q, \mathcal{A}) \in \mathbf{WC} \cap \mathbf{C_R}$. Moreover, the algorithm $\mathtt{pre}(q, \mathcal{A})$ is only able to increase prices and it never decreases them regardless of any buyer strategy.*

Proof. First, for each node $\mathfrak{n} \in \mathfrak{T}(\mathtt{pre}(q, \mathcal{A}))$, the recursion in Eq. (1) implies that there exists a node $\mathfrak{m} \in \mathfrak{T}(\mathcal{A})$ s.t. $\mathfrak{L}(\mathfrak{n}) \cong \mathtt{pre}\big(p^{n}, \mathfrak{L}(\mathfrak{m})\big)$ and $\mathfrak{R}(\mathfrak{n}) \cong \mathtt{pre}\big(p^{\mathfrak{m}}, \mathfrak{R}(\mathfrak{m})\big)$. In particular, $p^{\mathfrak{r}(\mathfrak{n})} = p^{\mathfrak{m}}$, $p^{\mathfrak{l}(\mathfrak{n})} = p^{\mathfrak{n}}$, $\wp(\mathfrak{L}(\mathfrak{n})) = \wp(\mathfrak{L}(\mathfrak{m})) \cup \{p^{\mathfrak{n}}\}$, and $\wp(\mathfrak{R}(\mathfrak{n})) = \wp(\mathfrak{R}(\mathfrak{m})) \cup \{p^{\mathfrak{m}}\}$. Let us prove by induction that $p^{\mathfrak{n}} \le p \; \forall p \in \wp(\mathfrak{T}(\mathfrak{n}))$. The basis of the induction: this condition is satisfied by the root node $\mathfrak{e}(\mathfrak{T}(\mathtt{pre}(q, \mathcal{A})))$ due to the choice of q. The inductive step: assume a node $\mathfrak{n} \in \mathfrak{T}(\mathtt{pre}(q, \mathcal{A}))$ satisfies this condition, then, for $p^{\mathfrak{l}(\mathfrak{n})}$ and $p^{\mathfrak{r}(\mathfrak{n})}$, we have $p^{\mathfrak{l}(\mathfrak{n})} = p^{\mathfrak{n}} \le p \; \forall p \in \wp(\mathfrak{L}(\mathfrak{n})) \subseteq \wp(\mathfrak{T}(\mathfrak{n}))$ and $p^{\mathfrak{r}(\mathfrak{n})} = p^{\mathfrak{m}} \le p \; \forall p \in \wp(\mathfrak{R}(\mathfrak{n})) = \wp(\mathfrak{R}(\mathfrak{m})) \cup \{p^{\mathfrak{m}}\}$, where we used $p^{\mathfrak{m}} \le p \; \forall p \in \wp(\mathfrak{R}(\mathfrak{m}))$ since the algorithm $\mathcal{A}$ is right-consistent.

Second, note that $p^{\mathfrak{n}} \le p \; \forall p \in \wp(\mathfrak{T}(\mathfrak{n})) \supseteq \wp(\mathfrak{R}(\mathfrak{n}))$, i.e., the definition of a right-consistent algorithm holds. Therefore, $\mathtt{pre}(q, \mathcal{A}) \in \mathbf{C_R}$ and the right-side part of weak consistency holds as well. Third, $p^{\mathfrak{l}(\mathfrak{n})} = p^{\mathfrak{n}} \forall \mathfrak{n} \in \mathfrak{T}(\mathtt{pre}(q, \mathcal{A}))$ as we noted above, and, hence, the left-side part of weak consistency is satisfied (the case of $p^{\mathfrak{l}(\mathfrak{n})} \neq p^{\mathfrak{n}}$ in Definition 2 of $\mathbf{WC}$ is never realized). Thus, $\mathtt{pre}(q, \mathcal{A}) \in \mathbf{WC}$ and it never decreases prices along any buyer strategy. □

4. Weakly Consistent Optimal Pricing

Let us apply the transformation $\mathtt{pre}$ to the pricing algorithm PRRFES (Drutsa, 2017) and refer to the transformed one as *prePRRFES*. Formally, the algorithm prePRRFES works in phases initialized by the phase index $l := 0$, the first offered price at the current phase $q_0 := 0$, and the iteration parameter $\epsilon_0 := 1/2$. At each phase $l \in \mathbb{Z}_+$, it sequentially offers prices $p_{l,k} := q_l + k\epsilon_l$, $k \in \mathbb{Z}_+$ (exploration; in contrast to PRRFES, it starts from $k = 0$), where $\epsilon_l := \epsilon_{l-1}^2 = 2^{-2^l}$, $l \in \mathbb{N}$. When a price $p_{l,k}$ is rejected, setting $K_l := k \ge 0$: (1) the algorithm offers this price p_{l,K_l} for $r - 1 \in \mathbb{Z}_+$ penalization rounds (if one of them is accepted, prePRRFES continues offering $p_{l,k}, k = K_l + 1, ..$ following Definition 4), (2) it offers the price p_{l,K_l} for $g(l) \in \mathbb{Z}_+$ exploitation rounds (buyer decisions made at them do not affect further pricing), and (3) prePRRFES goes to the next phase by setting $q_{l+1} := p_{l,K_l}$ and $l := l + 1$. The pseudo-code of prePRRFES is presented in Alg. B.1 (in Suppl.Mat.). The algorithm actions described above and code lines in Alg. B.1 that differ from the ones of PRRFES are highlighted in blue. Since PRRFES is a right-consistent algorithm, see (Drutsa, 2017), Lemma 1 implies that prePRRFES is both right-consistent and weakly consistent one.

Further in this section, we will show that the learning algorithm prePRRFES being properly configured is, in fact, a no-regret pricing and, moreover, is optimal with tight strategic regret bound of $\Theta(\log\log T)$. To show this, we adapt the methodology used to establishing the optimality of the algorithm PRRFES. Despite the PRRFES and its transformation prePRRFES look similar, this adaptation is not straightforward and, in particular, requires additional guarantees on acceptance of prices during the exploitation rounds.

Following (Mohri & Munoz, 2014; Drutsa, 2017) we assume that the discounting is geometric $\boldsymbol{\gamma} = \{\gamma^{t-1}\}_{t=1}^{\infty}$ from here on in this subsection. First, let us consider an analogue of (Drutsa, 2017, Prop.2) that will be useful to upper bound the strategic regret of the algorithm prePRRFES.

Proposition 1. *Let $\boldsymbol{\gamma} = \{\gamma^{t-1}\}_{t=1}^{\infty}$ be a discount sequence with $\gamma \in (0, 1)$, $\mathcal{A}$ be a pricing algorithm, $\mathfrak{n} \in \mathfrak{T}(\mathcal{A})$ be a starting node in a r-length penalization sequence (see Def. 4), all prices after r rejections are no lower than $p^{\mathfrak{n}}$ (i.e., $p^{\mathfrak{n}} \le p^{\mathfrak{m}} \; \forall \mathfrak{m} \in \mathfrak{L}(\mathfrak{l}^{r-1}(\mathfrak{n}))$), $t^{\mathfrak{n}} < T$, and $r > \log_{\gamma}(1-\gamma^2)$. If the price $p^{\mathfrak{n}}$ offered by the algorithm $\mathcal{A}$ at the node $\mathfrak{n}$ is rejected by the strategic buyer, then the following inequality on his valuation v holds:*

$$v - p^{\mathfrak{r}(\mathfrak{n})} < \eta_{r,\gamma}(p^{\mathfrak{r}(\mathfrak{n})} - p^{\mathfrak{n}}), \text{ where } \eta_{r,\gamma} := \frac{\gamma^r + \gamma - 1}{1 - \gamma^2 - \gamma^r}. \quad (2)$$

Proof sketch. Since the price $p^{\mathfrak{n}}$ is rejected, the following inequality holds, see (Mohri & Munoz, 2014, Lemma 1),

$$\gamma^{t^{\mathfrak{n}}-1}(v-p^{\mathfrak{n}})+S(\mathfrak{r}(\mathfrak{n}))<S(\mathfrak{l}^{r}(\mathfrak{n})), \qquad (3)$$

where $S(\mathfrak{m})$ denotes the surplus obtained by the buyer when playing an optimal strategy after reaching a node $\mathfrak{m}\in\mathfrak{T}(\mathcal{A})$. The left subtree's surplus $S(\mathfrak{l}^{r}(\mathfrak{n}))$ can be upper bounded as follows (using $p^{\mathfrak{n}}\leq p^{\mathfrak{m}}\ \ \forall\mathfrak{m}\in\mathfrak{L}(\mathfrak{l}^{r-1}(\mathfrak{n}))$):

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

while, in contrast to the proof of (Drutsa, 2017, Prop.2), the right subtree's surplus $S(\mathfrak{r}(\mathfrak{n}))$ is lower bounded by $\gamma^{t^{\mathfrak{n}}}(v-p^{\mathfrak{r}(\mathfrak{n})})$, because, after accepting $p^{\mathfrak{n}}$ at the round $t^{\mathfrak{n}}$, the buyer is able to earn at least this amount at the round $t^{\mathfrak{n}}+1$. We plug these bounds in Eq. (3), divide by $\gamma^{t^{\mathfrak{n}}-1}$, and obtain Eq. (2) after rearrangement. □

The full proof is given in Appendix A.1.1 in Supp.Mat. Note that the inequality in Eq. (2) bounds the deviation of the buyer's valuation v from the price offered at some node $\mathfrak{r}(\mathfrak{n})$ by some increment $p^{\mathfrak{r}(\mathfrak{n})}-p^{\mathfrak{n}}$. But, in contrast to (Drutsa, 2017, Prop.2), this bounding occurs when the buyer rejects the price $p^{\mathfrak{n}}$ offered previously to the price $p^{\mathfrak{r}(\mathfrak{n})}$ which is used as the reference price of the valuation's deviation.

As we show in the proof of Theorem 1, Prop. 1 allows us to obtain an upper bound for the number of exploring steps at each phase of the algorithm prePRRFES. However, this is *not enough to directly apply* the methodology of the proofs of (Drutsa, 2017, Th.3, Th.5) to bound the strategic regret, because, in contrast to the PRRFES, during exploitation rounds, the algorithm prePRRFES offers the price p_{l,K_l} *that has not been earlier accepted by the strategic buyer* (hence, there is no evidence to guarantee his acceptance during the exploitation). Namely, since the buyer's decision a_t made at an exploitation round t does not affect the algorithm's pricing in the subsequent rounds $t'>t$, the strategic buyer acts truthfully at this round t, i.e., $a_t=\mathbb{I}_{\{p_t\leq v\}}$. For the PRRFES, we knew that the price p_t was accepted in a previous round $t''<t$ (or $p_t=0$), but, for the prePRRFES, one has to specially guarantee the acceptance of the price p_t at the exploitation round t in the following proposition.

Proposition 2. *Let $\boldsymbol{\gamma}=\{\gamma^{t-1}\}_{t=1}^{\infty}$ be a discount sequence with $\gamma\in(0,1)$, $\mathcal{A}$ be a pricing algorithm, and $\mathfrak{n}\in\mathfrak{T}(\mathcal{A})$ be a starting node in a r-length penalization sequence (see Def. 4), which is followed by G exploitation rounds offering the price $p^{\mathfrak{n}}$ starting from the node $\mathfrak{l}^{r}(\mathfrak{n})$. If $r<\log_{\gamma}(1-\gamma)$, $G>\log_{\gamma}\big(1-(1-\gamma)\gamma^{-r}\big)$, $T\geq t^{\mathfrak{n}}+r+G-1$, and the buyer valuation v is higher than $p^{\mathfrak{n}}$ and lower than any price in the right subtree $\mathfrak{R}(\mathfrak{n})$ of the node $\mathfrak{n}$, i.e., $v<p\ \forall p\in\wp(\mathfrak{R}(\mathfrak{n}))$, then the strategic buyer rejects the price $p^{\mathfrak{n}}$ at the round $t^{\mathfrak{n}}$.*

Proof sketch. The condition $v<p\ \forall p\in\wp(\mathfrak{R}(\mathfrak{n}))$ implies that $S(\mathfrak{r}(\mathfrak{n}))=0$ and the strategic buyer will thus gain exactly $\gamma^{t^{\mathfrak{n}}-1}(v-p^{n})$ if he accepts the price p^{n} at the round $t^{\mathfrak{n}}$ ($S(\cdot)$ is defined in the proof of Prop. 1). Let us show that there exists a strategy in $\mathfrak{L}(\mathfrak{n})$ with a larger surplus. Indeed, if the buyer rejects r times the price $p^{\mathfrak{n}}$ and accepts this price G times after that, then he gets the following surplus:

$$\sum_{s=t^{\mathfrak{n}}+r}^{t^{\mathfrak{n}}+r+G-1}\gamma^{s-1}(v-p^{n})=(v-p^{n})\frac{1-\gamma^{G}}{1-\gamma}\gamma^{t^{n}-1+r}>(v-p^{n})\gamma^{t^{\mathfrak{n}}-1},$$

here the last inequality holds due to the condition on G. □

The proof's details are presented in Appendix A.1.2 in Suppl.Mat. Additionally to the claim of Proposition 2, note that, from the definitions of penalization and exploitation rounds, it follows that, if the strategic buyer rejects the price $p^{\mathfrak{n}}$ at the round $t^{\mathfrak{n}}$, he rejects this price $p^{\mathfrak{n}}$ at the rounds $t^{\mathfrak{n}}+1,\dots,t^{\mathfrak{n}}+r-1$ as well and accepts it at the rounds $t^{\mathfrak{n}}+r,\dots,t^{\mathfrak{n}}+r+G-1$. Note that, since $r\geq1$ (otherwise, there is no node $\mathfrak{n}$ and the right subtree $\mathfrak{R}(\mathfrak{n})$), the condition $r<\log_{\gamma}(1-\gamma)$ makes Prop. 2 meaningful only in the case of $\gamma>1/2$. This is consistent with a clear intuition that, having $\gamma\leq1/2$, the discount γ^{t-1} at a round t is no lower than the sum of all discounts in all possible subsequent rounds $\gamma^{t}/(1-\gamma)$, and the strategic buyer prefers thus to purchase a good for a price p_t at the t-th round, rather than many goods for a no lower price in all subsequent rounds.

In order to apply both Prop. 1 and Prop. 2, the number of penalization rounds r is required to be in the interval $I_{\gamma}=\big(\log_{\gamma}(1-\gamma^{2}),\log_{\gamma}(1-\gamma)\big)$. It easy to see that, for $\gamma\in(1/2,(\sqrt{5}-1)/2)$, this interval I_{γ} contains $r=1$ (i.e., there is no penalization). For $\gamma\in((\sqrt{5}-1)/2,1)$, the length of the interval I_{γ} is larger than 1, which guarantees thus existence of a natural number in I_{γ} (since $r\in\mathbb{N}$). For such discount rates, the following lemma provides values for r and G s.t. Prop. 1 and Prop. 2 hold and $\eta_{r,\gamma}$ (from Eq. (2)) is bounded by some positive number $\varkappa>0$. The proof is rather technical and is thus deferred to Appendix A.1.3.

Lemma 2. *Let $\widehat{\gamma}:=(\sqrt{5}-1)/2$, (a) for a discount rate $\gamma\in(\widehat{\gamma},1)$ and a constant $\varkappa>(1-\gamma)/(\gamma^{2}+\gamma-1)$, define*

$$r_{\gamma,\varkappa}:=\log_{\gamma}\left((1-\gamma)\Big(1+\frac{\varkappa}{1+\varkappa}\gamma\Big)\right)\quad \textit{and}$$

$$G_{\gamma,\varkappa}:=\log_{\gamma}\left(1-\Big(1+\frac{\varkappa}{1+\varkappa}\gamma\Big)^{-1}\gamma^{-1}\right);$$

(b) for $\gamma\in(1/2,\widehat{\gamma})$ and a constant $\varkappa\geq(2\gamma-1)/(1-\gamma^{2}-\gamma)$, define $r_{\gamma,\varkappa}:=1$ and $G_{\gamma,\varkappa}:=\log_{\gamma}(2\gamma-1)$. Then setting the number of penalization rounds $r=\lceil r_{\gamma,\varkappa}\rceil$ and the number of exploitation rounds $G\geq G_{\gamma,\varkappa}$ implies the conditions of Prop. 1 and 2 as well as $\eta_{r,\gamma}\leq\varkappa$.

In the interesting case of $\gamma=\widehat{\gamma}$, the interval I_{γ} is equal to $(1,2)$ and does not thus contain any natural number. This case is the point in which, on the one hand, we should apply penalization (i.e., set $r \geq 2$) according to Prop. 1 to control the amount of lies of the buyer, but, on the other hand, setting $r \geq 2$ results in the absence of a strongly more profitable buyer strategy in the left subtree of the node $\mathfrak{n}$ in Prop.2 than to accept a price at this node. We do not known whether the case of $\gamma = \widehat{\gamma}$ is a fundamental point in which prePRRFES could not be configured to avoid a linear regret, or it is a cause of insufficient power of our analytical tools. Further study of this case is left for future work. So, now we ready to obtain an upper bound for the prePRRFES for the case $\gamma \in (1/2, 1) \setminus \{\widehat{\gamma}\}$ by proving the following theorem.

Theorem 1. *Let $\boldsymbol{\gamma} = \{\gamma^{t-1}\}_{t=1}^{\infty}$ be a discount sequence with $\gamma \in (1/2,1)\setminus\{(\sqrt{5}-1)/2\}$ and let a constant $\varkappa$ satisfy the lower bounds from Lemma 2, where the constants $r_{\gamma,\varkappa}$ and $G_{\gamma,\varkappa}$ are defined as well. If $\mathcal{A}$ is the pricing algorithm prePRRFES with $r = \lceil r_{\gamma,\varkappa} \rceil$ and the exploitation rate $g(l) = \max\{2^{2^l}, \lceil G_{\gamma,\varkappa} \rceil\}, l \in \mathbb{Z}_+$, then, for any valuation $v \in [0,1]$ and $T \geq 2$, the strategic regret is upper bounded*[7]*:*

$$\begin{aligned} \mathrm{SReg}(T, \mathcal{A}, v, \boldsymbol{\gamma}) &\leq C(\log_2 \log_2 T + 2) + \frac{\lceil G_{\gamma,\varkappa} \rceil}{2} - 1, \\ C &:= rv + \frac{(1+\varkappa)}{2}(2 + \max\{2, \lceil G_{\gamma,\varkappa} \rceil\} + \varkappa). \end{aligned} \tag{4}$$

Proof sketch. Decompose the total regret over T rounds into the sum of the regrets during each phase: $\mathrm{SReg}(T, \mathcal{A}, v, \boldsymbol{\gamma}) = \sum_{l=0}^{L} R_l$, where L is the number of phases during T rounds. For the regret R_l at each phase except the last one, i.e., $l = 0, \ldots, L-1$, we have:

$$R_l = \sum_{k=0}^{K_l - 1} (v - p_{l,k}) + rv + g(l)(v - p_{l,K_l}), \tag{5}$$

where the first, second, and third terms correspond to the exploration rounds with acceptance, the reject-penalization rounds, and the exploitation rounds, respectively. First, here, we directly used Prop. 2 (via Lemma 2 since $g(l) \geq G_{\gamma,\varkappa}$) to conclude that $p_{l,K_l} < v$ and the price p_{l,K_l} is thus accepted by the strategic buyer at the exploitation rounds.

Second, rejection of the price p_{l,K_l} implies $v - p_{l,K_l+1} < \varkappa\epsilon_l$ by Prop. 1 via Lemma 2 since $\eta_{r,\gamma} \leq \varkappa$ for $r \geq \lceil r_{\gamma,\varkappa} \rceil$ and $\forall t \in \mathbb{N}$. Hence, the valuation $v \in \big(p_{l,K_l}, p_{l,K_l} + (1+\varkappa)\epsilon_l\big)$ and all accepted prices $p_{l+1,k}$, $\forall k \leq K_{l+1}$, from the next phase $l+1$ satisfy:

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

since any accepted price has to be lower than the valuation v for the strategic buyer. This infers $K_{l+1} < (1+\varkappa)\epsilon_l/\epsilon_{l+1} = (1+\varkappa)2^{2^l}$. Therefore, after some algebra, we have:

$$\begin{aligned} R_l &\leq \frac{(1+\varkappa)(2+\varkappa)}{2} + rv + g(l) \cdot (1+\varkappa)\epsilon_l \\ &\leq rv + \frac{(1+\varkappa)}{2}(2 + \max\{2, \lceil G_{\gamma,\varkappa} \rceil\} + \varkappa) \end{aligned} \tag{6}$$

for $l = 0, \ldots, L-1$, where we used the definition of the exploitation rate $g(l)$ to get $g(l) \cdot \epsilon_l \leq \max\{1, \lceil G_{\gamma,\varkappa} \rceil/2\}$.

The L-th phase differs from the other ones only in possible absence of some rounds: reject-penalization or exploitation ones. In this phase, we consider two cases on the actual number of exploitation rounds $g_L(L)$: (a) $g_L(L) \geq \lceil G_{\gamma,\varkappa} \rceil$ and (b) $g_L(L) < \lceil G_{\gamma,\varkappa} \rceil$. In the case (a), as above, we apply Prop. 2 to get that $p_{L,K_L} < v$ and the price p_{L,K_L} is thus accepted by the buyer at the exploitation rounds. This infer that R_L is upper bounded by the right-hand side of Eq. (5) with $l = L$ and, in turn, by the right-hand side of Eq. (6). In the case (b), we have no guarantee that $p_{L,K_L} < v$ and, therefore, p_{L,K_L} may be rejected by the strategic buyer at the exploitation rounds. Hence, we bound the regret R_L more roughly by $\frac{(1+\varkappa)(2+\varkappa)}{2} + (r + \lceil G_{\gamma,\varkappa} \rceil - 1)v$ and use the inequality $(\lceil G_{\gamma,\varkappa} \rceil - 1)v - \max\{1, \lceil G_{\gamma,\varkappa} \rceil/2\} \leq \lceil G_{\gamma,\varkappa} \rceil/2 - 1$. Finally, all our steps results in the bound:

$$\mathrm{SReg}(T, \mathcal{A}, v, \boldsymbol{\gamma}) = \sum_{l=0}^{L} R_l \leq C(L+1) + \frac{\lceil G_{\gamma,\varkappa} \rceil}{2} - 1$$

with C from Eq. (4). To obtain Eq. (4), get $L \leq \log_2 \log_2 T + 1$ via the same technique as in (Drutsa, 2017). □

The detailed proof is given in Appendix A.1.4 in Suppl.Mat. This theorem shows that *the learning algorithm prePRRFES is asymptotically optimal* since it has a tight strategic regret bound of $\Theta(\log\log T)$. At the same time, Theorem 1 *closes the open research question on the existence of a no-regret algorithm in the class* **WC** *(Drutsa, 2017)* since prePRRFES is weakly consistent. Note that the parameter $\varkappa$ can be adjusted either to minimize the number of penalization rounds r or to optimize the constant factor C making the upper regret bound finer. An attentive reader may also note that the pricing prePRRFES has the following drawback: this algorithm being applied against a myopic (truthful) buyer will incur a linear regret (in contrast to the source PRRFES). But we feel that this is the price we have to pay in order to construct a horizon-independent optimal algorithm that offers prices in a consistent manner (i.e., never revises prices that was previously reduced as it did by the PRRFES).

On the other hand, it is important to emphasize that our weakly consistent algorithm has the property: it never decreases prices (see Lemma 1). This may seem too drastic to obtain a no-regret algorithm from **WC**. But, in the next section, we show that this property, in fact, is nearly necessary for a weakly consistent algorithm to be a no-regret one.

[7]Note that the seller is not required to know γ precisely: if the seller knows only that $\gamma \in [\gamma' - \delta, \gamma' + \delta]$, then it is easy to get an upper bound similar to Eq.(4) by properly selecting parameters G and $\varkappa$ for some $\delta > 0$, when $\gamma' \in (1/2,1)\setminus\{(\sqrt{5}-1)/2\}$.

5. Weakly Consistent Algorithms with Double Decrease of Prices

In this section, we show that if a weakly consistent algorithm is able to doubly decrease prices, then it has a linear regret.

Definition 6. A weakly consistent algorithm $\mathcal{A}$ is said to be with *double decrease of prices* ($\mathcal{A}$ in the class $\mathbf{WC_{dd}}$) if there exists a path $\tilde{\mathbf{a}}$ in the tree $\mathfrak{T}(\mathcal{A})$ with the corresponding price sequence $\{\tilde{p}_t\}_{t=1}^{\infty}$ such that

$$\exists \tilde{t}_0, \tilde{t}_1 \in \mathbb{N} : \tilde{t}_0 \le \tilde{t}_1 \text{ and } \tilde{p}_{\tilde{t}_1+1} < \tilde{p}_{\tilde{t}_0} < p^{\mathfrak{e}(\mathfrak{T}(\mathcal{A}))}. \quad (7)$$

Theorem 2. *Let $\boldsymbol{\gamma} = \{\gamma_t\}_{t=1}^{\infty}$ be a discount sequence and $\mathcal{A} \in \mathbf{WC_{dd}}$ be a horizon-independent weakly consistent pricing algorithm with double decrease of prices and with the first offered price $p^{\mathfrak{e}(\mathfrak{T}(\mathcal{A}))} \in (0,1)$. Then there exists a valuation $v \in [0,1]$ s.t. $\mathrm{SReg}(T,\mathcal{A},v,\boldsymbol{\gamma}) = \Omega(T)$.*

Proof. We denote the first offered price as $p_1 := p^{\mathfrak{e}(\mathfrak{T}(\mathcal{A}))}$[8] and let the strategy $\tilde{\mathbf{a}}$, $\tilde{t}_0$, and $\tilde{t}_1$ be from Eq. (7). Let us decompose the set of all paths in the tree $\mathfrak{T}(\mathcal{A})$ into three sets $S_= \sqcup S_< \sqcup S_>$: (a) $S_=$ consists of paths whose price sequences $\{p_t\}_{t=1}^{\infty}$ are constant (i.e., $p_t{=}p_1\ \forall t{\in}\mathbb{N}$); (b) the price sequence $\{p_t\}_{t=1}^{\infty}$ of a path from $S_<$ ($S_>$, resp.) has the form: $\exists t_0{\in}\mathbb{N}$ s.t. $p_{t_0+1}{<}p_{t_0}$ ($p_{t_0+1}{>}p_{t_0}$, respectively) and $p_t{=}p_1, t{=}1,..,t_0$. The set $S_<$ is non-empty since $\tilde{\mathbf{a}}{\in}S_<$.

We consider the path $\hat{\mathbf{a}}$ s.t. $\hat{\mathbf{a}}_t{:=}\tilde{\mathbf{a}}_t\ \forall t{\le}\tilde{t}_1$ and $\hat{\mathbf{a}}_t{:=}1\ \forall t{>}\tilde{t}_1$ (it coincides with $\tilde{\mathbf{a}}$ at up to the $\tilde{t}_1$-th round) and its corresponding price sequence $\{\hat{p}_t\}_{t=1}^{\infty}$. Note that $\hat{\mathbf{a}} \in S_<$. Let us denote $\Delta = p_1 - \hat{p}_{\tilde{t}_0} > 0$, then, due to the weak consistency of the algorithm $\mathcal{A}$[9], we have $\hat{p}_t \le \hat{p}_{\tilde{t}_0} = p_1 - \Delta,\ \forall t \ge \tilde{t}_1$. Consider a buyer with a valuation $v_\varepsilon := p_1 + \varepsilon, \varepsilon > 0$. If the buyer follows the strategy $\hat{\mathbf{a}}$, then his surplus can be lower bounded:

$$\mathrm{Sur}_{\boldsymbol{\gamma}}(T,\mathcal{A},v_\varepsilon,\hat{\mathbf{a}}) \ge \textstyle\sum_{t=\tilde{t}_1+1}^{T} \gamma_t(\Delta+\varepsilon) \quad \forall T > \tilde{t}_1. \quad (8)$$

If this buyer follows any strategy $\mathbf{a}$ from $S_>$, then one has

$$\mathrm{Sur}_{\boldsymbol{\gamma}}(T,\mathcal{A},v_\varepsilon,\mathbf{a}) \le \textstyle\sum_{t=1}^{T} \gamma_t\varepsilon \quad \forall \mathbf{a} \in S_>\ \forall T > 0, \quad (9)$$

because the price sequence corresponding to $\mathbf{a}$ satisfies $p_t \ge p_1\ \forall t \in \mathbb{N}$. Let $\varepsilon_0 := \min\left\{\Delta \cdot \gamma_{\tilde{t}_1+1} / \sum_{t=1}^{\tilde{t}_1} \gamma_t, 1 - p_1\right\}$, then, $\forall \varepsilon \in (0,\varepsilon_0)$ we have $v_\varepsilon \in (0,1)$ and

$$\varepsilon < \Delta \cdot \textstyle\sum_{t=\tilde{t}_1+1}^{T} \gamma_t / \sum_{t=1}^{\tilde{t}_1} \gamma_t \qquad \forall T > \tilde{t}_1.$$

Hence, the right-hand side of Eq. (8) is larger than the one of Eq. (9), which implies that $\mathrm{Sur}_{\boldsymbol{\gamma}}(T,\mathcal{A},v_\varepsilon,\mathbf{a}) < \mathrm{Sur}_{\boldsymbol{\gamma}}(T,\mathcal{A},v_\varepsilon,\hat{\mathbf{a}})\ \ \forall \mathbf{a} \in S_>$.

Thus, we showed that, for $T > \tilde{t}_1$, there exists a strategy in $S_<$ (i.e. $\hat{\mathbf{a}}$) that is better (in terms of discounted surplus) than any strategy in $S_>$ for the strategic buyer with any valuation $v_\varepsilon = p_1 + \varepsilon$ s.t. $\varepsilon \in (0,\varepsilon_0)$. Hence, the optimal strategy $\mathbf{a}^{\mathrm{Opt}} \in S_= \cup S_<$ for $T > \tilde{t}_1$ and, for any strategy $\mathbf{a}$ from this union, the regret $\mathrm{Reg}(T,\mathcal{A},v_\varepsilon,\mathbf{a})$ is lower bounded by $\sum_{t:a_t=0} v_\varepsilon + \sum_{t:a_t=1}(v_\varepsilon - p_1) \ge T\varepsilon$. This implies the bound for the strategic regret: $\mathrm{SReg}(T,\mathcal{A},v_\varepsilon,\boldsymbol{\gamma}) \ge T\varepsilon = \Omega(T), T > \tilde{t}_1$, since $\tilde{t}_1$ and ε are independent of T. □

Remark 1. For simplicity, in Def. 6 and Th. 2, we considered the root node $\mathfrak{e}(\mathfrak{T}(\mathcal{A}))$ from which the double decrease of prices should start. In fact, we can replace it both in Def. 6 and in Th. 2 (without harm for the proof) by any node $\mathfrak{n} \in \mathfrak{T}(\mathcal{A})$ s.t. it is passed by the strategic buyer.

Note that this theorem holds for any discount sequence and has thus the following corollary that generalizes (Drutsa, 2017, Th.4) to any discounting (the proof is in App. A.2.1).

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

The key intuition behind Theorem 2 consists in the following: the strategic buyer can lie few times to decrease offered prices and, due to (even weak) consistency, receive prices at least on $\varepsilon > 0$ lower than his valuation v all the remaining rounds. Note that the buyer is able to mislead a wide range of weakly consistent algorithms: the set $\mathbf{WC_{dd}}$ is significantly larger than the set $\mathbf{RWC}$ (previously known as the largest subset of $\mathbf{WC}$ with a linear regret). On the other hand, if we do not allow the buyer to apply this intuition, one can build an algorithm with a sublinear strategic regret, as we showed in Section 4.

6. Conclusions

We studied the scenario of repeated posted-price auctions with a strategic buyer that holds a fixed private valuation and discounts his cumulative surplus, while the seller applies a horizon-independent online learning (discrete) algorithm to set prices. First, we closed the open research question on the existence of a no-regret horizon-independent weakly consistent algorithm by proposing a novel algorithm that never decreases offered prices and can be applied against strategic buyers with a tight strategic regret bound of $\Theta(\log\log T)$. Second, we showed that the property of non-decreasing prices is crucial and, in fact, is nearly necessary for a weakly consistent algorithm to be a no-regret one. Finally, we introduced non-trivial techniques such as (a) a novel transformation which maps a right-consistent algorithm to a both weakly and right- consistent one that is only able to increase prices; (b) approaches to control and guarantee acceptance of exploitation prices that have not been earlier accepted.

[8] Note that p_1 is the first element in a price sequence of any buyer strategy for a given $\mathcal{A}$.

[9] It easy to derive (using the weak consistency) that $\hat{p}_t \le \hat{p}_{\tilde{t}_0}\ \forall t{\ge}\tilde{t}_0$. Because, otherwise, if $\exists t' : \tilde{t}_0{<}t'{<}\tilde{t}_1$ s.t. $\hat{p}_{t'}{>}\hat{p}_{\tilde{t}_0}$, which implies $\hat{p}_t{\ge}\hat{p}_{\tilde{t}_0}\ \forall t{>}t'$ and contradicts to $\hat{p}_{\tilde{t}_1}{<}\hat{p}_{\tilde{t}_0}$.

Acknowledgements

I thank Gleb Gusev for helpful discussions on the presentation of this work and anonymous reviewers for useful suggestions to improve the paper.

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Chawla, S., Devanur, N. R., Karlin, A. R., and Sivan, B. Simple pricing schemes for consumers with evolving values. In *SODA'2016*, pp. 1476–1490, 2016.

Chen, X. and Wang, Z. Bayesian dynamic learning and pricing with strategic customers. *SSRN*, 2016.

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

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

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

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

Devanur, N. R., Peres, Y., and Sivan, B. Perfect bayesian equilibria in repeated sales. In *SODA'2015*, pp. 983–1002, 2015.

Drutsa, A. Horizon-independent optimal pricing in repeated auctions with truthful and strategic buyers. In *WWW'2017*, pp. 33–42, 2017.

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

Feldman, M., Koren, T., Livni, R., Mansour, Y., and Zohar, A. Online pricing with strategic and patient buyers. In *NIPS'2016*, pp. 3864–3872, 2016.

Fudenberg, D. and Villas-Boas, J. M. Behavior-based price discrimination and customer recognition. *Handbook on economics and information systems*, 1:377–436, 2006.

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

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

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

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

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

Immorlica, N., Lucier, B., Pountourakis, E., and Taggart, S. Repeated sales with multiple strategic buyers. In *EC'2017*, pp. 167–168, 2017.

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

Kanoria, Y. and Nazerzadeh, H. Dynamic reserve prices for repeated auctions: Learning from bids. 2014.

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

Krishna, V. *Auction theory*. Academic press, 2009.

Lahaie, S., Medina, A. M., Sivan, B., and Vassilvitskii, S. Testing incentive compatibility in display ad auctions. In *WWW'2018*, 2018.

Leme, R. P., Syrgkanis, V., and Tardos, É. Sequential auctions and externalities. In *SODA'2012*, pp. 869–886. SIAM, 2012.

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

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

Medina, A. M. and Vassilvitskii, S. Revenue optimization with approximate bid predictions. In *NIPS'2017*, pp. 1856–1864, 2017.

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

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

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

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

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

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

Nisan, N., Roughgarden, T., Tardos, E., and Vazirani, V. V. *Algorithmic game theory*, volume 1. v.1 CUPC, 2007.

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

Paes Leme, R., Pál, M., and Vassilvitskii, S. A field guide to personalized reserve prices. In *WWW'2016*, 2016.

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

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

Schmidt, K. M. Commitment through incomplete information in a simple repeated bargaining game. *Journal of Economic Theory*, 60(1):114–139, 1993.

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

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

Vanunts, A. and Drutsa, A. Optimal pricing in repeated posted-price auctions. *arXiv preprint arXiv:1805.02574*, 2018.

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

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

Varian, H. R. and Harris, C. The vcg auction in theory and practice. *The A.E.R.*, 104(5):442–445, 2014.

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

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

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

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