Optimal Non-parametric Learning in Repeated Contextual Auctions with Strategic Buyer

Alexey Drutsa [1 2]

Abstract

We study learning algorithms that optimize revenue in repeated contextual posted-price auctions where a seller interacts with a single strategic buyer that seeks to maximize his cumulative discounted surplus. The buyer's valuation of a good is a fixed private function of a d-dimensional context (feature) vector that describes the good being sold. In contrast to existing studies on repeated contextual auctions with strategic buyer, in our work, the seller is not assumed to know the parametric model that underlies this valuation function. We introduce a novel non-parametric learning algorithm that is horizon-independent and has tight strategic regret upper bound of $\Theta(T^{d/(d+1)})$. We also non-trivially generalize several value-localization techniques of non-contextual repeated auctions to make them effective in the considered contextual non-parametric learning of the buyer valuation function.

1. Introduction

Real-time ad exchanges (RTB), search engines, and other Internet companies consider *revenue maximization* as one of the most important development directions (Gomes & Mirrokni, 2014; Balseiro et al., 2015; Agarwal et al., 2014; Charles et al., 2016; Drutsa, 2017b; 2018; Hummel, 2018). A large part of ad inventory is sold via widely applicable second-price auctions (He et al., 2013; Mohri & Medina, 2014) and its extensions as GSP (Varian, 2007; Sun et al., 2014) or VCG (Varian, 2009; Varian & Harris, 2014), while revenue in most of them is usually maximized by means of reserve prices (Myerson, 1981; Krishna, 2009; Cesa-Bianchi et al., 2013; Paes Leme et al., 2016; Drutsa, 2020). A large number of online auctions run, e.g., by RTB involve only a single advertiser (Amin et al., 2013; Mohri & Munoz, 2014; Drutsa, 2017b; 2018; Vanunts & Drutsa, 2019). In this case, a second-price auction with reserve reduces to a *posted-price auction* (Kleinberg & Leighton, 2003): the seller sets a reserve price for a good (e.g., an ad space) and the buyer (e.g., advertiser) decides whether to reject or accept this price (to bid below or above it).

We study a scenario when a single seller repeatedly interacts through a posted-price auction with the *same strategic* buyer that holds a *private valuation* for a good and *seeks to maximize his cumulative discounted surplus* (Amin et al., 2013; 2014). This strategic scenario was studied in the case when the buyer's valuation is constant over all rounds (Amin et al., 2013; Mohri & Munoz, 2014; Drutsa, 2017b; 2018; 2020). However, in practice of online advertising, an Ad exchange (the seller) faces with different ad spaces (goods) sequentially offered to the same advertiser (the buyer): e.g., different users look at different web pages whose ad spaces are offered for sale.

This more realistic scenario can be modeled as follows: *the buyer private valuation depends on the context (feature vector) that describes the good being sold*. In other words, this dependence is fixed and is unknown to the seller, while the context of a currently offered good is observed by both the buyer and the seller. In our work, this scenario with a fixed valuation function (of feature vectors) is studied. The seller uses an online learning algorithm, which is announced to the buyer in advance and, in each round, selects the price based on (a) previous decisions of the buyer and (b) the observed context information of the goods offered for sale up to the current round. The seller maximizes her cumulative revenue over a finite time horizon T via regret minimization, i.e., he seeks for a pricing algorithm with a sublinear regret on T (a no-regret pricing) (Amin et al., 2013; 2014; Mohri & Munoz, 2014; Drutsa, 2017b; 2018).

The main weakness of the existing algorithms (Amin et al., 2014; Golrezaei et al., 2019) in a scenario of repeated contextual auctions *with strategic buyer* is their assumption that the valuation function is a particular parametric model of features (namely, a linear model or a kernel one): these pricing algorithms are thus aimed to reveal the parameters of this model. However, in practice, it is very natural that the

[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 37^{th} International Conference on Machine Learning, Online, PMLR 119, 2020.

seller does not know in advance which parametric model is used by the buyer to derive the value of a good from its feature vector. The methods developed for parametric models cannot be effectively used in this case. Hence, in this paper, *we focus on the situation in which the seller does not know the buyer's parametric model of his valuation and wants to learn this valuation in a non-parametric way assuming it to be a Lipschitz function*[1]. To the best of our knowledge, no existing study investigated worst-case regret optimizing algorithms that set prices in repeated contextual auctions with the strategic buyer whose valuation is a fixed private Lipschitz function of a d-dimensional context vector.

In our study, we propose a novel horizon-independent optimal algorithm that can be applied against our strategic Lipschitz buyer with tight regret upper bound of $\Theta(T^{d/(d+1)})$ (Th. 1). This result constitutes *the main contribution of our work* and closes the open research question on the existence of a no-regret (and moreover, optimal) pricing for the setup of repeated contextual auctions. We also show that the approaches used in the fixed valuation (non-contextual) setup (Drutsa, 2017b) (namely, instruments to locate the valuation and to build a horizon-independent pricing) can be non-trivially upgraded to be successfully used in the contextual setup. So, construction and analysis of the proposed algorithm and the upgraded instruments have required introduction of novel techniques, which are contributed by our work as well. They include: (a) the method to isolate so-called penalization rounds (see Def. 2 and Prop. 1) and exploitation rounds (see, Sec. 5), what allows to converge algorithm prices to the valuation and to do it independently of the horizon; and (b) the guarantee on the amount of lie from the buyer when he accepts an offered price (see Prop. 2).

2. Setup of Repeated Contextual Auctions

We study the following mechanism of *repeated contextual posted-price auctions*. Namely, a *single seller* repeatedly proposes goods (e.g., ad opportunities) to a *single buyer* over T rounds (*the time horizon*): one good per round. The good proposed in a round t is represented by a d-dimensional feature vector $x_t \in \mathbb{X} := [0,1]^d$ also referred to as the *context* of the round, $d \in \mathbb{Z}_+$. The buyer holds a *private valuation* $v(x) \in [0,1]$ for a good with context $x \in \mathbb{X}$: the valuation function $v : [0,1]^d \to [0,1]$ is unknown to the seller and does not depend on the rounds. This valuation function v is assumed to be L-Lipschitz on $\mathbb{X}$, $L > 0$, where the class of d-dimensional L-Lipschitz functions is defined by $\mathrm{Lip}_L(\mathbb{X}) := \{f : \mathbb{X} \to [0,1] \mid \forall y, z \in \mathbb{X}\ |f(y) - f(z)| \le L\|y - z\|_\infty\}$[2] as in (Mao et al., 2018). So, for each round t, (a) the good's feature vector $x_t \in \mathbb{X}$ is observed both by the seller, and by the buyer; (b) a price $p_t \in \mathbb{R}_+$ is offered by the seller; and (c) the buyer (knowing x_t and p_t) makes an allocation decision $a_t \in \{0,1\}$: $a_t = 1$, when he accepts to buy the currently offered good at that price, $a_t = 0$, otherwise.

In our setup, the seller applies a *pricing algorithm* $\mathcal{A}$ that sets prices $\{p_t\}_{t=1}^T$ in response to buyer decisions $a_{1:T} := \{a_t\}_{t=1}^T$ and observed feature vectors $x_{1:T} := \{x_t\}_{t=1}^T$. We consider the deterministic online learning scenario when the price p_t in a round $t \in \{1, ..., T\}$ can depend only on the buyer's actions $a_{1:t-1}$[3] during the previous rounds and the observed context information $x_{1:t}$ up to the current round. Following (Drutsa, 2017b; 2018), 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 her. Let $\mathbf{A}^d$ be the set of such algorithms.

Hence, given a pricing algorithm $\mathcal{A} \in \mathbf{A}^d$, buyer decisions $a_{1:T}$ and goods $x_{1:T}$ uniquely define the price sequence $p_{1:T} := \{p_t\}_{t=1}^T$, which infers the seller's total revenue $\sum_{t=1}^T a_t p_t$. This revenue is compared to the revenue that would have been earned by offering the buyer's valuations $\{v(x_t)\}_{t=1}^T$ if they were known in advance to the seller (Kleinberg & Leighton, 2003; Amin et al., 2013; 2014; Mohri & Munoz, 2014; Drutsa, 2017b; 2018). This leads to the definition of the *regret* of the algorithm $\mathcal{A}$ that faced the buyer with the valuation function $v : [0,1]^d \to [0,1]$ making decisions $a_{1:T}$ for the goods $x_{1:T}$ over T rounds as $\mathtt{Regret}(T, \mathcal{A}, v, a_{1:T}, x_{1:T}) := \sum_{t=1}^T (v(x_t) - a_t p_t)$.

As in (Amin et al., 2013; 2014; Mohri & Munoz, 2014; Drutsa, 2017b; 2018; Golrezaei et al., 2019), we assume that the seller's algorithm $\mathcal{A}$ is announced to the buyer *in advance*. The buyer can then act strategically against this algorithm: in each round t, the buyer makes the optimal allocation decision $a_t = a_t^{\mathrm{Opt}}(T, \mathcal{A}, v, \gamma, a_{1:t-1}, x_{1:t}, D)$ that maximizes his expected future γ-discounted surplus $\mathbb{E}_{x_s \sim D}[\sum_{s=t}^T \gamma^{s-1} a_s (v(x_s) - p_s)]$, where $\gamma \in (0,1]$ is the buyer's discount rate and D is a probability distribution over the feature domain $\mathbb{X}$ for goods x_s, $s \ge t+1$, (Amin et al., 2014). Note that the distribution D is only used by the buyer to estimate future goods' features (for rounds $\ge t+1$) and to make thus the strategic decision. The seller is not required to know this distribution: our results hold for any sequence of goods $x_{1:T}$ (see the optimization goal below and Th. 1). When T rounds has been played with goods $x_{1:T}$, we can define *the strategic regret* of the algorithm $\mathcal{A}$ that faced the strategic buyer with the valuation function $v : [0,1]^d \to [0,1]$ over T rounds as $\mathtt{SRegret}(T, \mathcal{A}, v, \gamma, x_{1:T}, D) :=$

[1] Lipschitz assumption is required since the seller can get a linear regret against non-Lipschitz valuation even against a truthful buyer (Mao et al., 2018). See more discussion on this in Sec 6.

[2] $\|z\|_\infty := \max_i |z^i|$, $z = (z^1, \ldots, z^d)$, is the ℓ_∞-norm on $\mathbb{R}^d$. Our results hold for other ℓ_p-norms as well.

[3] $y_{t_1:t_2} := \{y_t\}_{t=t_1}^{t_2}$ denotes a part of a time series $\{y_t\}_{t=1}^T$.

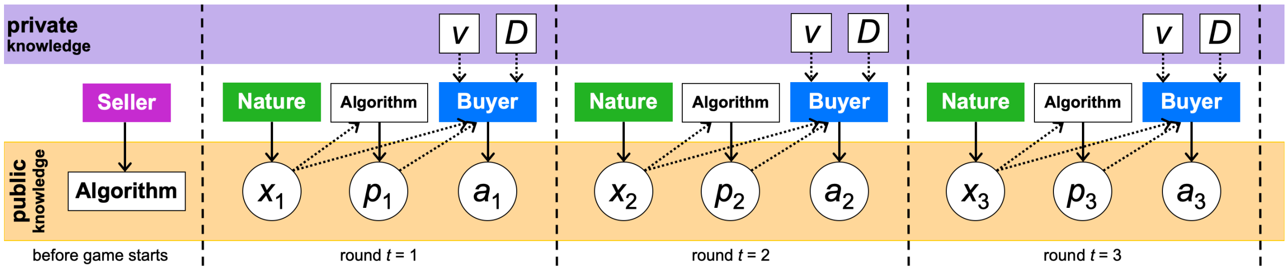

Figure 1. The game's workflow (an algorithm and the buyer exploit public knowledge available in the previous rounds).

$\texttt{Regret}\big(T, \mathcal{A}, v, a^{\mathrm{Opt}}_{1:T}(T, \mathcal{A}, v, \gamma, x_{1:T}, D), x_{1:T}\big)$.

We are interested in pricing algorithms that have $o(T)$ strategic regret for the *worst-case* valuation function $v \in \mathrm{Lip}_L(\mathbb{X})$ and the *worst-case* sequence of goods $x_{1:T} \in \mathbb{X}^T$. Formally, an algorithm $\mathcal{A}$ is said to be a *no-regret* one when $\sup_{x_{1:T}\in\mathbb{X}^T, v\in\mathrm{Lip}_L(\mathbb{X}), D} \texttt{SRegret}(T, \mathcal{A}, v, \gamma, x_{1:T}, D) = o(T)$. The optimization goal is to find an algorithm with the best possible asymptotic of the form $O_{T\to\infty}(f(T))$.

Note that our no-regret requirements are much stronger than the ones of (Amin et al., 2014; Golrezaei et al., 2019): e.g., our regret bounds will hold for any realization of the context series $x_{1:T}$ independently of the distribution D used by the buyer. E.g., this distribution D can be treated as the buyer's personal beliefs about future goods. Also note that all our results hold if this distribution D changes between rounds. In Figure 1, we depict the game's workflow for an illustration of the information structure and roles of the seller, the buyer, and the nature (adversary).

3. Related Work and Background

Related Work. There are two lines of works that are the most relevant to ours. The first one (Amin et al., 2014; Cohen et al., 2016; Golrezaei et al., 2019; Leme & Schneider, 2018; Mao et al., 2018; Javanmard & Nazerzadeh, 2019; Javanmard et al., 2019; Zhiyanov & Drutsa, 2020) dealt with contextual pricing in repeated auctions (or multidimensional search in online fashion). The works (Cohen et al., 2016; Leme & Schneider, 2018; Mao et al., 2018; Javanmard & Nazerzadeh, 2019; Javanmard et al., 2019) assumed that the buyer's behavior is myopic (truthful) in a round, while our study considers the seller's interactions with a strategic buyer that optimizes his cumulative surplus. The algorithms of (Cohen et al., 2016; Leme & Schneider, 2018; Javanmard & Nazerzadeh, 2019; Javanmard et al., 2019; Zhiyanov & Drutsa, 2020) search a valuation function as a parametric model, while the ones of (Mao et al., 2018) do it in a non-parametric way. Key results of the latter study are discussed in our overview of pricing algorithms below.

The second line of works studied our strategic setup with fixed private valuation, but in the non-contextual case (all goods are equal) (Amin et al., 2013; Mohri & Munoz, 2014; Drutsa, 2017b;a; 2018; 2020; Schmidt, 1993; Hart & Tirole, 1988; Devanur et al., 2015; Immorlica et al., 2017; Vanunts & Drutsa, 2019). We overview the most interesting and relevant results from these works in the next subsection as well. The studies (Amin et al., 2014; Golrezaei et al., 2019; Zhiyanov & Drutsa, 2020) lie at the intersection of both lines of works: their authors considered contextual repeated auctions where the seller interacted with the same strategic buyer. In contrast to our work, their algorithms explicitly assume that the valuation function has a particular parametric model, what makes them inapplicable in our more realistic scenario when this knowledge is unavailable to the seller in advance (see Sec. 1).

Background on Pricing Algorithms. First, our scenario in the absence of context information for goods (i.e., when $d = 0$) reduces to the setup of repeated posted-price auctions earlier introduced in (Amin et al., 2013). In this case, the strategic buyer has a fixed private valuation for all goods, and pricing algorithms for worst-case regret optimization were well studied (Amin et al., 2013; Mohri & Munoz, 2014; Drutsa, 2017b; 2018). In particular, it is known that, if the buyer cumulative utility is not discounted over rounds (i.e., the discount rate $\gamma = 1$), there does not exist a no-regret pricing, i.e., the linear strategic regret has lower bound of $\Omega(T)$ (Amin et al., 2013). Since, in our setup, the features are chosen adversarially, this lower bound holds in the studied repeated contextual auctions as well.

For the other cases $\gamma \in (0, 1)$, the lower bound of $\Omega(\log\log T)$ holds (Kleinberg & Leighton, 2003; Mohri & Munoz, 2014), and two optimal algorithms with tight strategic regret bound of $\Theta(\log\log T)$ have been recently proposed for the non-contextual setup (Drutsa, 2017b; 2018). Their construction strongly relied on the technique of penalization (Mohri & Munoz, 2014; Drutsa, 2017b). These results has been extended to a setting with multiple buyers in a round (Drutsa, 2020), where a special transformation div was applied to 1-buyer pricing to obtain an optimal algorithm with tight strategic regret bound of $\Theta(\log\log T)$ in repeated second-price auctions. To the best of our knowledge, our study is the first one where this approach is exploited for a contextual setup. Moreover, we show that the penalization technique without a significant modification could not be effectively applied here (see Sec. 4).

Second, the recent work (Mao et al., 2018) studied a quite similar setup of repeated contextual auctions with a buyer that held a fixed Lipschitz valuation. However, their scenario considered the buyer that made decisions myopically (truthfully) in each round (formally, their case is covered by ours when $\gamma = 0$). Two optimal algorithms with tight truthful regret bound of $\Theta(T^{d/(d+1)})$ were proposed (Mao et al., 2018). One of them relied on the idea of "iterative partition refinement" widely used in learning with Lipschitz conditions (Kleinberg, 2005; Hazan & Megiddo, 2007; Kleinberg et al., 2008; Slivkins, 2014; Mao et al., 2018). Our algorithm exploits this idea as well, but we emphasize that the truthful and the strategic scenarios are fundamentally different. The algorithms designed to act against a myopic buyer cannot be thus straightforwardly used in the strategic setting as it was shown for the non-contextual setup (Drutsa, 2017b) and as we discuss in Sec. 4: special tools for localization of the buyer valuation are required (otherwise, the strategic buyer may mislead the algorithm). So, *our research goal comprises closing of the open research question on the existence of a no-regret algorithm for the considered scenario of repeated contextual auctions with strategic Lipschitz buyer*.

4. Search Algorithms

4.1. Auxiliary Definitions

In order to parameterize all possible prices that can be offered by an algorithm $\mathcal{A} \in \mathbf{A}^d$, one introduces the notion of a *state of the game* that encodes a history of passed rounds. Formally, $\mathfrak{S}_t := \mathbb{X}^t \times \{0,1\}^{t-1}$ is the set of all possible states after receiving the good in the t-th round (before the buyer's decision). Let $\mathfrak{S} := \sqcup_{t=1}^{\infty}\mathfrak{S}_t$ be *the set of states* in all rounds. Thus, any algorithm $\mathcal{A} \in \mathbf{A}^d$ can be bijectively associated with a map $\mathrm{p} : \mathfrak{S} \to \mathbb{R}_+$, i.e., it maps any observed history $\mathfrak{s} = (x_{1:t}, a_{1:t-1}) \in \mathfrak{S}_t$ to the price $p_t = \mathrm{p}(\mathfrak{s})$. In other words, $\mathrm{p}(x_{1:t}, a_{1:t-1})$ is the price that will be offered by this algorithm to the buyer in response to the feature vector x_t in the t-th round after $t-1$ rounds with the feature vectors $x_{1:t-1}$ and the buyer decisions $a_{1:t-1}$. For any state $(x_{1:t}, a_{1:t-1}) \in \mathfrak{S}$, let the sets

$$\mathfrak{R}(x_{1:t}, a_{1:t-1}) := \{(z_{1:s}, b_{1:s-1}) \in \mathfrak{S} \mid s > t, z_{1:t} = x_{1:t}, b_{1:t-1} = a_{1:t-1}, b_t = 1\}$$

and

$$\mathfrak{L}(x_{1:t}, a_{1:t-1}) := \{(z_{1:s}, b_{1:s-1}) \in \mathfrak{S} \mid s > t, z_{1:t} = x_{1:t}, b_{1:t-1} = a_{1:t-1}, b_t = 0\}$$

be all possible states that continue the history $(x_{1:t}, a_{1:t-1})$ when the buyer accepts and rejects, respectively, the price in the round t. The set $\mathfrak{R}(\mathfrak{s})$ (the set $\mathfrak{L}(\mathfrak{s})$) is referred to as *the right (left, respectively) continuations* of a state $\mathfrak{s} \in \mathfrak{S}$[4].

We also use the following notations for subsets of states $(x_{1:t}, a_{1:t-1})$ whose *last feature vector* x_t belongs to a set $X \subseteq \mathbb{X}$: $\mathfrak{S}_t(X) := \mathbb{X}^{t-1} \times X \times \{0,1\}^{t-1}$, $\mathfrak{S}(X) := \sqcup_{t=1}^{\infty}\mathfrak{S}_t(X)$, $\mathfrak{R}(\mathfrak{s}; X) := \mathfrak{R}(\mathfrak{s}) \cap \mathfrak{S}(X)$, and $\mathfrak{L}(\mathfrak{s}; X) := \mathfrak{L}(\mathfrak{s}) \cap \mathfrak{S}(X)$ for $\mathfrak{s} \in \mathfrak{S}$. In other words, a state $\tilde{\mathfrak{s}} \in \mathfrak{R}(\mathfrak{s}; X)$ (or $\in \mathfrak{L}(\mathfrak{s}; X)$) is such right (left, resp.) continuation of the state $\mathfrak{s}$ that its current context vector belongs to the set X.

For any subset $X \subseteq \mathbb{X}$, let $\mathrm{diam}(X) := \sup_{x,y \in X} \|x - y\|_\infty$ denote the ℓ_∞-diameter of this set.

4.2. Notion of a Search Algorithm

A good algorithm for finding (learning) the valuation function $v(\cdot)$ intuitively should work as follows: it keeps track of two functions $u, w : \mathbb{X} \to [0,1]$; in each round t with a feature vector x_t, it proposes a price $p_t \in [u(x_t), w(x_t)]$; and, after the buyer's decision, the algorithm can increase the values of $u(\cdot)$ and can decrease the ones of $w(\cdot)$ (in some points of $\mathbb{X}$). We refer to such pricing as a *search algorithm*. The key idea behind this algorithm is, first, to locate the valuation function v between u and w (i.e., $u(x) \le v(x) \le w(x)\ \forall x \in \mathbb{X}$) and, second, to tight the gap between u and w (i.e., "$w(x) - u(x) \to 0$ as $t \to \infty$" $\forall x \in \mathbb{X}$). In the non-contextual setup (i.e., $d = 0$), the most known examples of search algorithms are the binary search and its generalization — a consistent algorithm (Mohri & Munoz, 2014; Drutsa, 2017b) — that tracks a feasible search interval $[q, q']$ and reduces it to either $[q, p]$ or $[p, q']$ depending on the buyer's response to a price p. The "midpoint" algorithm and similar ones from (Mao et al., 2018) are fresh examples of the ones with search behavior described above.

Definition 1. Given any algorithm $\mathcal{A} \in \mathbf{A}^d$, for any state $\mathfrak{s} \in \mathfrak{S}$ and any point $x \in \mathbb{X}$, we introduce *the lower bound of future prices* $u_{\mathfrak{s},1}(x) := \inf_{\mathfrak{s}' \in \mathfrak{R}(\mathfrak{s};\{x\})} \mathrm{p}(\mathfrak{s}')$ when the buyer accepts the price $\mathrm{p}(\mathfrak{s})$ and the one $u_{\mathfrak{s},0}(x) := \inf_{\mathfrak{s}' \in \mathfrak{L}(\mathfrak{s};\{x\})} \mathrm{p}(\mathfrak{s}')$ when he rejects it. *The upper bounds of future prices* are defined similarly: $w_{\mathfrak{s},1}(x) := \sup_{\mathfrak{s}' \in \mathfrak{R}(\mathfrak{s};\{x\})} \mathrm{p}(\mathfrak{s}')$ and $w_{\mathfrak{s},0}(x) := \sup_{\mathfrak{s}' \in \mathfrak{L}(\mathfrak{s};\{x\})} \mathrm{p}(\mathfrak{s}')$.

It is easy to see that $u(\cdot)$ and $w(\cdot)$ in Def. 1 behave as in the above description of search algorithms: $p_t \in [u(x_t), w(x_t)]$, values of $u(\cdot)$ can only increase, while values of $w(\cdot)$ can only decrease. However, this functions u and w cannot guarantee the correct location of the valuation ($u(x) \le v(x) \le w(x)\ \forall x \in \mathbb{X}$) as well as tightening of the gap (i.e., "$w(x) - u(x) \to 0$ as $t \to \infty$" $\forall x \in \mathbb{X}$). To provide guarantees on correct location, we will introduce novel tools presented in Prop. 1 and 2 further in this section; while guarantees on

[4]Note that the set of states $\mathfrak{S}$ is a generalization of the complete binary tree $\mathfrak{T}$ which is actively used in studies on the non-contextual repeated auctions (Mohri & Munoz, 2014; Drutsa, 2017b; 2018). Moreover, when $d = 0$, the set of states $\mathfrak{S}$ is exactly the tree $\mathfrak{T}$, any state $\mathfrak{s} \in \mathfrak{S}$ is a node in this tree, while $\mathfrak{R}(\mathfrak{s})$ and $\mathfrak{L}(\mathfrak{s})$ are the right and left subtrees of the node $\mathfrak{s}$, respectively. So, the state representation of contextual algorithms $\mathbf{A}^d$ is important as the tree representation of non-contextual algorithms $\mathbf{A}^0$.

tightening of the gap (and strategic regret as well) will be given for our novel algorithm in Lemma 2 of Sec. 5.

4.3. Localization of Buyer Valuation

In the non-contextual repeated auction setup, the strategic buyer may mislead any consistent algorithm used by the seller and cause thus a linear regret (Drutsa, 2017b, Th.4). Since the features are chosen adversarially in our contextual setting, if a pricing algorithm from $\mathbf{A}^d$ behaves as a consistent one within one point in $\mathbb{X}$ (e.g., "midpoint" algorithm and its analogue for pricing loss (Mao et al., 2018)), then this pricing will also have a linear regret by the results of (Drutsa, 2017b). More generally, if an algorithm makes decisions relying on the assumption "rejection (acceptance) of p_t for a good x_t implies $v(x_t) \le p_t$ ($v(x_t) \ge p_t$, resp.)" as in (Mao et al., 2018), then the seller may obtain a linear regret interacting with the strategic buyer. In order to reveal information on the buyer valuation from his binary signal a_t in a round t, the special tool (Drutsa, 2017b, Prop.2) has been found in the non-contextual setting; it is based on so-called penalization rounds (Mohri & Munoz, 2014; Drutsa, 2017b). We non-trivially extend the notion of penalization to make it useful in our contextual setup (inapplicability of the original notion is discussed after Prop. 1).

Definition 2. For a pricing algorithm $\mathcal{A} \in \mathbf{A}^d$, a state $\mathfrak{s} \in \mathfrak{S}_\tau$ (and the corresponding round τ) is referred to as a *penalization* one, if its price $\mathrm{p}(\mathfrak{s}) = 1$ and, for all right continuations, we have: $\mathrm{p}(\mathfrak{s}') = 1 \ \forall \mathfrak{s}' \in \mathfrak{R}(\mathfrak{s})$. A state $\tilde{\mathfrak{s}} \in \mathfrak{S}_t$ is said to be the *start of a r-length penalization with domain* $\tilde{X} \subseteq \mathbb{X}$, if any left continuation $(x_{1:s}, a_{1:s-1}) \in \mathfrak{L}(\tilde{\mathfrak{s}}; \tilde{X})$ s.t. at most its $r-1$ future context vectors x_{t_1}, ..., x_{t_k} belongs to $\tilde{X}$ after the t-th round (i.e., $k \le r-1$ and $t < t_j, j \le k$), is a penalization state.

Informally, in the state $\tilde{\mathfrak{s}}$, after rejection of the price in the t-th round, the algorithm $\mathcal{A}$ will conduct penalization in future $r-1$ rounds whose feature vectors belongs to the penalization domain $\tilde{X}$[5] (i.e., penalization affects $r-1$ goods only from $\tilde{X}$). The strategic buyer will never accept the price in a penalization round since, otherwise, the price of any of future goods (even outside $\tilde{X}$) will be 1, which is at least his valuation.

The algorithm that will be proposed in Sec. 5 partitions the whole context set $\mathbb{X}$ into subsets (domains) and conducts pricing actions independently between these subsets. In order to formalize this independence we introduce the following notion of isolation.

[5]Consider an example: $d=1$, $\tilde{X}=[0.5,1]$, $r=3$. Let $x_1=0.7$ be a start of 3-length penalization with domain $[0.5,1]$ and the buyer rejects the price in $t=1$. The penalization rounds (for this penalization start) will be those whose context sequence $x_{1:s}$ is, e.g., $(0.7, 0.6)$, $(0.7, 0.2, 0.9)$, $(0.7, 0.2, 0.4, 0.7)$, but *not* $(0.7, 0.2)$ and *not* $(0.7, 0.2, 0.4)$. See Appendix E for more details.

Definition 3. For a pricing algorithm $\mathcal{A} \in \mathbf{A}^d$ and a state $\mathfrak{s} = (x_{1:t}, a_{1:t-1}) \in \mathfrak{S}$, a set of contexts $X \subseteq \mathbb{X}$ is said to be *isolated (from $\mathbb{X} \setminus X$) after the state $\mathfrak{s}$ on* if $x_t \in X$ and buyer decisions made for any future good in the set X (i.e., in any state $\{\mathfrak{s}\} \cup \mathfrak{L}(\mathfrak{s}, X) \cup \mathfrak{R}(\mathfrak{s}, X)$) do not affect the algorithm prices for any future goods whose feature vectors are in $\mathbb{X} \setminus X$ (i.e., prices in any state $\mathfrak{L}(\mathfrak{s}, \mathbb{X} \setminus X) \cup \mathfrak{R}(\mathfrak{s}, \mathbb{X} \setminus X)$).

So, for an algorithm that conducts isolated pricing in some set of feature vectors, the following contextual analogue of (Drutsa, 2017b, Prop.2) can be proved.

Proposition 1. *Let $\gamma \in (0,1)$, $\mathcal{A}$ be a pricing algorithm, the state $\tilde{\mathfrak{s}} = (x_{1:t}, a_{1:t-1}) \in \mathfrak{S}$ be the start of a r-length penalization with domain $\tilde{X} \subseteq \mathbb{X}$ (see Def. 2), and the set $\tilde{X}$ be isolated from $\mathbb{X} \setminus \tilde{X}$ after the state $\tilde{\mathfrak{s}}$ on (see Def. 3), where $r > \log_\gamma(1-\gamma)$ and $x_t \in \tilde{X}$. Then, if the price $p(\tilde{\mathfrak{s}})$ is rejected by the strategic buyer with a L-Lipschitz valuation function $v \in \mathrm{Lip}_L(\mathbb{X})$, then the following inequality on his valuation $v(x_t)$ holds:*

$$v(x_t) - p(\tilde{\mathfrak{s}}) < \zeta_{r,\gamma}\big(p(\tilde{\mathfrak{s}}) + L\mathrm{diam}(\tilde{X}) - \inf_{x \in \tilde{X}} u_{\tilde{\mathfrak{s}},0}(x)\big) \quad \text{where} \quad \zeta_{r,\gamma} := \gamma^r/(1-\gamma-\gamma^r). \tag{1}$$

Proof. Since the buyer's decisions for goods in $\tilde{X}$ (starting from this round t) do not affect the algorithm's prices for future goods with feature vectors outside $\tilde{X}$, one can analyze buyer surplus only in rounds with goods from $\tilde{X}$. So, let $S_{\tilde{X}}(\sigma)$ denote the buyer's expected future surplus over goods from $\tilde{X}$ when he follows a strategy σ. Let σ^{Opt} be the optimal strategy of the buyer in the round t. Let σ' be the strategy, where the buyer accepts the good in the round t and rejects each future good from $\tilde{X}$ (σ' coincides with σ^{Opt} for goods in $\mathbb{X} \setminus \tilde{X}$). Since σ^{Opt} is optimal, this implies that $S_{\tilde{X}}(\sigma') \le S_{\tilde{X}}(\sigma^{\mathrm{Opt}})$ (see Appendix A.1 for details). By definition $S_{\tilde{X}}(\sigma') = \gamma^{t-1}(v(x_t) - p(\tilde{\mathfrak{s}}))$, while the right-hand side of the inequality can be upper bounded as follows

$$S_{\tilde{X}}(\sigma^{\mathrm{Opt}}) = \mathbb{E}[\sum_{s=t}^{T} \gamma^{s-1} a_s (v(x_s) - p_s) \mid x_s \in \tilde{X}, \sigma^{\mathrm{Opt}}] \le$$
$$\le \sum_{s=t+r}^{T} \gamma^{s-1} \sup_{x \in \tilde{X}} [v(x) - u_{\tilde{\mathfrak{s}},0}(x)],$$

where we (1) used that the first $r-1$ rounds with goods in $\tilde{X}$ will be penalization ones (the buyer will certainly reject them), and (2) upper bounded instant surpluses in all further rounds as maximal possible ones by $\sup_{x \in \tilde{X}}(v(x) - u_{\tilde{\mathfrak{s}},0}(x))$ (see Def. 1 for $u_{\tilde{\mathfrak{s}},0}(\cdot)$). The latter expression can be trivially bounded by $v(x_t) + L\mathrm{diam}(\tilde{X}) - \inf_{x \in \tilde{X}} u_{\tilde{\mathfrak{s}},0}(x)$. Combining all inequalities together, one obtains: $(v(x_t) - p(\mathfrak{s}))\gamma^{t-1} \le \big[v(x_t) + L\mathrm{diam}(\tilde{X}) - \inf_{x \in \tilde{X}} u_{\tilde{\mathfrak{s}},0}(x)\big]\gamma^{t-1+r}/(1-\gamma)$, what implies Eq. (1) after dividing by γ^{t-1} and rearrangement of terms, since $r > \log_\gamma(1-\gamma)$. $\square$

Thus, Prop. 1 is a tool to locate the valuation $v(x_t)$ despite the strategic buyer makes decisions non-myopically. Moreover, since $v \in \mathrm{Lip}_L(\mathbb{X})$, Eq. (1) can by used to upper bound the valuation of any good: $v(x) - p(\tilde{\mathfrak{s}}) < \zeta_{r,\gamma}(p(\tilde{\mathfrak{s}}) + L\mathrm{diam}(\tilde{X}) - \inf_{z \in \tilde{X}} u_{\tilde{\mathfrak{s}},0}(z)) + L\|x_t - x\|_\infty \ \forall x \in \mathbb{X}$. So, for any context $x \in \tilde{X}$, we get

$$v(x) < p(\tilde{\mathfrak{s}}) + \zeta_{r,\gamma}(p(\tilde{\mathfrak{s}}) - \inf_{z \in \tilde{X}} u_{\tilde{\mathfrak{s}},0}(z)) + (1+\zeta_{r,\gamma})L\mathrm{diam}(\tilde{X}).$$

We see that the closeness of this bound to the rejected price $p(\tilde{\mathfrak{s}})$ depends, first, on the distance of this price to the lower bound of future prices (as in the non-contextual case (Drutsa, 2017b)) and, second, on the size of the domain $\tilde{X}$. Hence, tightening of the gap between the lower bound u and the upper bound w is not possible here if the size of the isolated domain $\tilde{X}$ does not decrease as well. *It is the key contrast to the non-contextual setting* (where it is enough to offer a price closer to the last accepted one) and *the reason why the original penalization (Drutsa, 2017b) is useless*: the original penalization penalizes the next $r-1$ goods independently of their context (the case $\tilde{X} = \mathbb{X}$ in Def. 2), what keeps the gap between the bounds u and w at least L.

In Prop. 1, we generalized the non-contextual instrument (Drutsa, 2017b, Prop.2) to control possible lies from the buyer when he rejects a price. What about lies when the buyer accepts a price? In the non-contextual scenario, usage of a right-consistent algorithm guarantees that the strategic buyer has no incentive to lie during price acceptance, because all future prices will be at least the accepted price (Drutsa, 2017b). In our contextual setup with Lipschitz valuation, the information on v revealed in a current point x_t needs to be used to get information on v in neighbor points of x_t and, thus, future prices may be lower than the accepted one. So, the following guarantee on possible buyer lie during price acceptance is needed:

Proposition 2. *Let $\gamma \in (0,1)$, $\mathcal{A}$ be a pricing algorithm, the state $\mathfrak{s} = (x_{1:t}, a_{1:t-1}) \in \mathfrak{S}$ be s.t. the feature vector $x_t \in \tilde{X} \subseteq \mathbb{X}$ and $p(\mathfrak{s}) - \inf_{x \in \tilde{X}} u_{\mathfrak{s},1}(x) \geq (1+\gamma)L\mathrm{diam}(\tilde{X})/(1-\gamma)$, while the context set $\tilde{X}$ is isolated from $\mathbb{X} \setminus \tilde{X}$ after $\mathfrak{s}$ on (see Def. 3). Then, if the price $p(\mathfrak{s})$ is accepted by the strategic buyer with Lipschitz valuation function $v \in \mathrm{Lip}_L(\mathbb{X})$, then the following inequality on his valuation $v(x_t)$ holds:*

$$v(x_t) \geq \inf_{x \in \tilde{X}} u_{\mathfrak{s},1}(x) + L\mathrm{diam}(\tilde{X}). \qquad (2)$$

Proof sketch. Analysis here is similar to the one of Prop. 1, but, in the case of acceptance, we can infer for the surplus within $\tilde{X}$: $S_{\tilde{X}}(\sigma^{\mathrm{Opt}}) \geq 0$, since the buyer can reject all goods in $\tilde{X}$ and get at least 0. Then, get the upper bound $S_{\tilde{X}}(\sigma^{\mathrm{Opt}}) \leq \gamma^{t-1}(v(x_t) - p_t) + \sum_{s=t+1}^{T} \gamma^{s-1} \sup_{x \in \tilde{X}}[v(x) - u_{\mathfrak{s},1}(x)]$. Finally, use the condition on $p(\mathfrak{s}) - \inf_{x \in \tilde{X}} u_{\mathfrak{s},1}(x)$ to obtain Eq. (2). Full proof is in Appendix A.2 in Supp.Materials. □

Similarly to Prop. 1, Eq. (2) can by used to lower bound the valuation of any good $x \in \mathbb{X}$: $v(x) \geq \inf_{z \in \tilde{X}} u_{\mathfrak{s},1}(z) + L\mathrm{diam}(\tilde{X}) - L\|x_t - x\|_\infty$. So, for any $x \in \tilde{X}$, we get $v(x) \geq \inf_{z \in \tilde{X}} u_{\mathfrak{s},1}(z)$, what is seemingly as in the non-contextual case (Drutsa, 2017b): the valuation is no lower than the lower bound for future prices. However, it is partially similar: rearrange the condition on $u_{\mathfrak{s},1}$ in Prop. 2: $\inf_{z \in \tilde{X}} u_{\mathfrak{s},1}(z) \leq p(\mathfrak{s}) - (1+\gamma)L\mathrm{diam}(\tilde{X})/(1-\gamma)$. We see that the valuation lower bound in Eq. (2) holds only if the future prices will be lower than the offered price by at least $(1+\gamma)L\mathrm{diam}(\tilde{X})/(1-\gamma)$. This margin grows as $\gamma \to 1$ and linearly depends on the size of the domain $\tilde{X}$. So, similarly to the discussion after Prop. 1, tightening of the gap between the lower bound u and the upper bound w is possible (via Prop. 2) only when the size of the isolated domain $\tilde{X}$ decreases as well. In the next section, we propose an algorithm that exploits the valuation localization tools (Prop. 1 and 2) and shrinks sizes of isolated parts of $\mathbb{X}$ when the localization of the valuation becomes more precise.

5. Penalized Exploiting Lipschitz Search

From here on we assume that $d \geq 1$ (i.e., the non-contextual case $d = 0$ is not considered). We introduce the *Penalized Exploiting Lipschitz Search* pricing algorithm (*PELS*) that will be shown to have tight strategic regret bound of $\Theta(T^{d/(d+1)})$ (see Th. 1). This pricing uses the penalization rounds (Mohri & Munoz, 2014; Drutsa, 2017b) upgraded in Def. 2 and ideas of "iterative partition refinement" (Kleinberg, 2005; Hazan & Megiddo, 2007; Kleinberg et al., 2008; Slivkins, 2014; Mao et al., 2018). First, PELS has three parameters: *the price offset* $\eta \in [1, +\infty)$, *the degree of penalization* $r \in \mathbb{N}$, and *the exploitation rate* $g : \mathbb{Z}_+ \to \mathbb{Z}_+$. Second, this algorithm keeps track of a partition $\mathcal{X}$ of the feature domain $\mathbb{X}$ initialized to $\lceil (4\eta + 6)L \rceil^d$ cubes (boxes) whose side length is $l = 1/\lceil (4\eta + 6)L \rceil$: $\mathcal{X} = \{[0, l], (l, 2l], \ldots, (1-l, 1]\}^d$. For each box $X \in \mathcal{X}$, PELS also keeps track of the lower bound $u^X \in [0,1]$, the upper bound $w^X \in [0,1]$, and the depth $m^X \in \mathbb{Z}_+$. At the beginning of the game, they are initialized as follows: $u^X = 0, w^X = 1$, and $m^X = 0, X \in \mathcal{X}$. Third, the workflow of the algorithm is organized independently in each box $X \in \mathcal{X}$. Namely, in a round t, the algorithm receives a good with a feature vector $x_t \in \mathbb{X}$ and finds the box $X \in \mathcal{X}$ in the current partition $\mathcal{X}$ such that $x_t \in X$. Then, the proposed price p_t is determined only from the current state associated with the box X, while the buyer decision a_t is used only to update the state associated with this box X.

So, in each box $X \in \mathcal{X}$, the algorithm iteratively offers exploration price $u^X + \eta L\mathrm{diam}(X)$. If this price is accepted by the buyer, the lower bound u^X is increased by $L\mathrm{diam}(X)$. If this price is rejected, the upper bound w^X is decreased by $(w^X - u^X) - 2(\eta+1)L\mathrm{diam}(X)$, and 1 is offered as a penalization price for $r-1$ next rounds in this box X (if one of them is accepted, PELS continues offering 1 all the remaining rounds, following Def. 2 with the penalization domain $\tilde{X}=X$). Finally, if, after an acceptance of an exploration price or after penalization rounds, we have $w^X - u^X < (2\eta+3)L\mathrm{diam}(X)$, then PELS:

1. offers the exploitation price u^X for $g(m^X)$ next rounds in this box X (buyer decisions made at them do not affect further pricing);
2. bisects each side of the box X to obtain 2^d boxes $\mathcal{X}_X := \{X_1, ..., X_{2^d}\}$ with ℓ_∞-diameter $\mathrm{diam}(X)/2$;
3. refines the partition $\mathcal{X}$ replacing the box X by the new boxes $\mathcal{X}_X$[6]. These new boxes $\mathcal{X}_X$ inherit the state of the lower bound u^X and the upper bound w^X from the current state of X, while their depth $m^{X'} = m^X + 1\ \forall X' \in \mathcal{X}_X$.

The pseudo-code of PELS is presented in Alg. B.1 in Appendix B of Supplementary Materials.

As we see from the description, the goal of the PELS parameters is as follows: (1) $g(\cdot)$ controls the number of exploitation rounds before we split a box into 2^d boxes; (2) r corresponds to the number of penalization rounds and is responsible to satisfy Prop. 1; and (3) η controls the position of exploration price in the interval $[u^X, w^X]$ and is responsible to satisfy Prop. 1 and 2. Further, exact values for these parameters will be provided to make this algorithm be no-regret in our strategic setting (see Eq. (3)).

Note that PELS, in fact, tracks functions u and $w : \mathbb{X} \to [0,1]$ defined by: $u \equiv \sum_{X\in\mathcal{X}} u^X \mathbb{I}_X$ and $v \equiv \sum_{X\in\mathcal{X}} v^X \mathbb{I}_X$ for a current state of PELS with partition $\mathcal{X}$, where $\mathbb{I}_X$ is the indicator[7] of a set $X \in \mathcal{X}$. These functions are initialized as $u \equiv 0$ and $w \equiv 1$. An attentive reader may note that, in each state of PELS, they match with the functions in Def. 1, if one does not take into account penalization rounds in calculation of $\sup$ (for w). In particular, the function u will be used to obtain statements of Prop. 1 and 2 for PELS.

5.1. Optimality of the PELS Algorithm

For $\gamma_0 \in (0,1)$, let the parameters of the PELS algorithm be set as follows:

$$r \geq r_{\gamma_0} := \left\lceil \log_{\gamma_0}\left((1-\gamma_0)/2\right)\right\rceil, \\ g(m) := 2^m\ \forall m \in \mathbb{Z}_+, \quad \text{and} \quad \eta \geq 2/(1-\gamma_0). \tag{3}$$

Then, for any $\gamma \in (0, \gamma_0]$, we have $\zeta_{r,\gamma} \leq 1$ in Eq. (1), and we will show that PELS is optimal with tight strategic regret of $\Theta(T^{d/(d+1)})$. First, let us show that the functions u and $w : \mathbb{X} \to [0,1]$ (tracked by PELS) correctly locate the buyer valuation function v.

Lemma 1. *In any state of the algorithm PELS with parameters in Eq. (3), the functions $u \equiv \sum_{X\in\mathcal{X}} u^X \mathbb{I}_X$ and $v \equiv \sum_{X\in\mathcal{X}} v^X \mathbb{I}_X$ correctly locate the valuation function v of the strategic buyer with discount $\gamma \in (0, \gamma_0]$, i.e., $\forall x \in \mathbb{X}$, the following inequalities hold: $u(x) \leq v(x) \leq w(x)$.*

Proof. One needs only to prove that if the inequalities hold before an exploration round, then they will still hold when either u or w is updated after. Let $\mathfrak{s} \in \mathfrak{S}_T$ be the current state of PELS, x_t be the feature vector in the current exploration round t, and $X \in \mathcal{X}$ be the current box ($x_t \in X$). Then the offered price is $p_t = u^X + \eta L\mathrm{diam}(X)$. If the buyer accepts p_t, then the updated lower bound is set to $\hat{u}^X = u^X + L\mathrm{diam}(X)$. Hence, using Eq. (3): $p_t - \hat{u}^X = (\eta-1)L\mathrm{diam}(X) \geq (1+\gamma)L\mathrm{diam}(X)/(1-\gamma)$. So, the conditions of Prop. 2 are satisfied by PELS in the current round since $\inf_{x\in X} u_{\mathfrak{s},1}(x) = \hat{u}^X$. Therefore, from Eq. (2) and $v \in \mathrm{Lip}_L(\mathbb{X})$, we have $v(x) \geq v(x_t) - L\mathrm{diam}(X) \geq \hat{u}^X\ \forall x \in X$. On the other hand, if the buyer rejects p_t, then the updated upper bound is set to $\hat{w}^X = u^X + 2(\eta+1)L\mathrm{diam}(X)$, the conditions of Prop. 1 are satisfied by PELS in the current round, and $\inf_{x\in X} u_{\mathfrak{s},0}(x) = u^X$. Therefore, from Eq. (1), Eq. (3), and $v \in \mathrm{Lip}_L(\mathbb{X})$ we have $v(x) \leq v(x_t) + L\mathrm{diam}(X) \leq 2p_t - u^X + 2L\mathrm{diam}(X) = \hat{w}^X\ \forall x \in X$. □

Second, we can show that the gap between the functions u and w tights with some guarantees, what is formalized in the following lemma (the proof is technical and is in App.A.3).

Lemma 2. *In any exploration round with $x_t \in X \in \mathcal{X}$ of the algorithm PELS with parameters in Eq. (3): (a) the gap $w^X - u^X$ reduces by at least $L\mathrm{diam}(X)$ and at most $(w^X - u^X) - 2(\eta+1)L\mathrm{diam}(X)$ after an update of either u or w; and (b) the difference between the functions u and v satisfies the following bounds: $(2\eta+3)L\mathrm{diam}(X) \leq w^X - u^X \leq \lceil(4\eta+6)L\rceil\mathrm{diam}(X)\ \forall X \in \mathcal{X}$.*

So, combining Lemmas 1 and 2, we can conclude that PELS learns the valuation function v. In order to get guarantees on the regret of this learning, we need one more statement.

Lemma 3. *In any box X of a depth $m \in \mathbb{Z}_+$, the algorithm PELS with parameters in Eq.(3) conducts at most $2\eta+3+L^{-1}$ exploration rounds, at most $(r-1)(2\eta+3+L^{-1})$ penalization ones, and at most 2^m exploitation rounds. The strategic regret in this box is at most $r(2\eta+3+L^{-1})+2^{-1}$.*

Proof. By Lemma 2, the gap $w^X - u^X$ is bounded and reduces at least by $L\mathrm{diam}(X)$ in each exploration round.

[6] Note that, after this procedure, $\mathcal{X}$ remains a partition of the feature domain $\mathbb{X}$, i.e., $\cup_{X\in\mathcal{X}} X = \mathbb{X}$.

[7] i.e., $\mathbb{I}_X(x) = 1$, when $x \in X$, and 0, otherwise

Hence, the number of exploration rounds is at most

$$\frac{\lceil(4\eta+6)L\rceil\mathrm{diam}(X)-(2\eta+3)L\mathrm{diam}(X)}{L\mathrm{diam}(X)}\le 2\eta+3+L^{-1}.$$

Each exploration round can result in up to $(r-1)$ penalization ones in the box X (by Def. 2). The number of exploitation rounds is controlled by $g(m)$ and is at most 2^m from Eq. (3). Each exploration round with acceptance contributes at most $w^X - u^X \le \lceil(4\eta+6)L\rceil\mathrm{diam}(X) = \lceil(4\eta+6)L\rceil 2^{-m}\lceil(4\eta+6)L\rceil^{-1} = 2^{-m}$ to the regret. Each exploration round with rejection and each penalization one contribute at most $w^X \le 1$. Each exploitation round has regret at most $w^X - u^X$ when it becomes less than $(2\eta+3)L\mathrm{diam}(X) = 2^{-m}(2\eta+3)L/\lceil(4\eta+6)L\rceil \le 2^{-m-1}$. □

Theorem 1. *Let $d \ge 1$, $\gamma_0 \in (0,1)$, and $\mathcal{A}$ be the PELS algorithm with parameters set as in Eq.(3). Then, for any discount $\gamma \in (0,\gamma_0]$, distribution D, and feature vectors $x_{1:T} \in \mathbb{X}^T$, the regret of PELS against the strategic buyer with a Lipschitz valuation $v \in \mathrm{Lip}_L(\mathbb{X})$ is upper bounded:*

$$\texttt{SRegret}(T,\mathcal{A},v,\gamma,x_{1:T},D)\le C\left(N_0(T+N_0)^d\right)^{\frac{1}{d+1}}, \quad (4)$$

where $C = 2^{d+1}r(2\eta+3+L^{-1})+1$, $N_0 = \lceil(4\eta+6)L\rceil^d$, and $\mathbb{X} = [0,1]^d$.

Proof. Assume that the game has been played. Let N_m denote the number of boxes with a depth $m \in \mathbb{Z}_+$ that belong(ed) to the partition $\mathcal{X}$ (including those that have been bisected). First, we know that $N_0 = \lceil(4\eta+6)L\rceil^d$ and $N_m \le 2^{md}N_0 \;\forall m \in \mathbb{Z}_+$. Second, in order to have a box of depth $m \in \mathbb{N}$, a box of depth $m-1$ has to be bisected and must have $g(m-1)$ exploitation rounds passed. So, the total number of rounds T is at least $\sum_{m>0} g(m-1)N_m$. Third, the upper bound on the total strategic regret is $R = \sum_m C(L,r,\eta)N_m$, where $C(L,r,\eta) := r(2\eta+3+L^{-1}) + 2^{-1}$ (see Lemma 3). Thus, we have to upper bound R that linearly depends on $\{N_m\}$ given the linear constraints $0 \le N_m \le 2^{md}N_0 \;\forall m$ and $\sum_{m>0} g(m-1)N_m \le T$ for a growing $g(\cdot)$.

Solving this linear program (see Appendix C.1), we find that R achieves the maximum, when $N_m = 2^{md}N_0 \;\forall m < M$ and $N_m = 0 \;\forall m > M$ for some $M \in \mathbb{N}$. Therefore, the bound on the total regret $R \le \sum_{m=0}^{M} C(L,r,\eta)2^{md}N_0 = C(L,r,\eta)N_0(2^{d(M+1)}-1)/(2^d-1)$. Finally, we bound 2^M as follows:

$$T \ge \sum_{m=1}^{M-1} g(m-1)N_m = \sum_{m=1}^{M-1} 2^m 2^{md}N_0 = N_0\frac{2^{(d+1)M}-2^{d+1}}{2^{d+1}-1},$$

what implies $2^M \le 2(1+T/N_0)^{1/(d+1)}$. Hence, $R \le 2^{d+1}C(L,r,\eta)N_0^{1/(d+1)}(T+N_0)^{d/(d+1)}$ and we get Eq. (4) after rearrangement. □

6. Extensions and Discussion

6.1. Analysis of the Regret Upper Bound

First, note that the upper bound in Eq. (4) on strategic regret is tight and of the form $\Theta(T^{d/(d+1)})$, since there exists a lower bound of $\Omega(T^{d/(d+1)})$ for any algorithm that is aimed to reveal Lipschitz function with pricing loss (Mao et al., 2018, Th.7). Hence, *PELS is optimal*. Second, it is also easy to see that the bound grows as $L \to \infty$, what coincides with the result for the myopic scenario (Mao et al., 2018) and the intuition that the larger L is the harder for the seller to propagate revealed information for the current good to its neighborhood in the d-dimensional space.

Third, the dependence of Eq. (4) on d shows that the larger the dimension of the feature space $\mathbb{X}$ is, the slower the growth of the seller's revenue. This situation could not be improved in our setting due to the lower bound (Mao et al., 2018): even in the truthful scenario, we have lower bound for regret of any algorithm and this bound has exponential dependence in d[8]. Finally, the bound in Eq. (4) blows up when $\gamma \to 1$ what is in the line with the same property of the optimal bound for non-contextual setting (Drutsa, 2017b; 2018) and with the case of $\gamma = 1$, where any algorithm has linear regret (Amin et al., 2013).

6.2. On Lipschitz Valuation

It is known that any absolutely continuous (or differentiable almost everywhere) function is a Lipschitz one. So, when a buyer uses a continuous parametric model for his valuation (e.g., linear one, kernel model, etc), then our algorithm can be used. Note that linear or kernel models are the most popular scenarios considered in related work (Amin et al., 2014; Cohen et al., 2016; Leme & Schneider, 2018; Javanmard & Nazerzadeh, 2019; Javanmard et al., 2019; Golrezaei et al., 2019), but, in our work, the seller does not need to know which kernel is used by the buyer (remind: the seller does not know which parametric model is used by the buyer). Therefore, we believe that our setup is more realistic than the ones studied before, since ours covers the models considered in those studies. For sure, it might be a case when the buyer utilizes a non-Lipschitz function for his valuation (e.g., with jump discontinuity points), but such a scenario is a good direction for future work. However, even in this case, our result suggests some insights: note that our upper bound in Eq. (4) and the lower bound in (Mao et al., 2018, Th.7) blow up when $L \to \infty$. Hence, without knowing the position of the discontinuity points, it seems hard (or even impossible) to find a no-regret algorithm, since the non-Lipschitz function can be approximated by a series of Lipschitz ones with growing $L \to \infty$.

[8]See (Mao et al., 2018, proof of Th.7 in Appendix), where the bound is estimated as α^d with $\alpha > 1$ if $T \to \infty$.

6.3. Horizon-Independence

The algorithm PELS is horizon-independent by its construction. Note that PELS can be used against a myopic buyer in the scenario "pricing loss" of (Mao et al., 2018): just set $r{=}1$ and $\eta{=}2$. Hence, *we provide a horizon-independent alternative to the horizon-dependent algorithms from (Mao et al., 2018)*. Note that those algorithms of Mao et al. (2018) can be made horizon-independent via the state-of-the-art "doubling" trick (Cesa-Bianchi et al., 2013; Heidari et al., 2016; Cohen et al., 2016): split the time line into epochs; run a horizon-dependent algorithm for an epoch; and, after the end of the epoch, restart the algorithm (from the initial learning state) but in a 2-time longer epoch. However, modified in this way algorithms have the weakness: they do no exploit the learned information from previous epochs, what may unnecessarily increase the regret (usually, it raises the constant factor in the regret (Drutsa, 2017b)). In contrast to this, PELS does not suffer from this drawback, since its exploration and exploitation rounds are interspersed neatly (without returning to the initial state), see Sec. 5.

6.4. On Efficient Implementation for Practice

The proposed algorithm PELS has exponentiality in d in terms of both computational complexity and memory requirements. In both directions one can use some tricks to make efficient implementation. First, the most computationally complex part of our algorithm is the search of the box X that contains the context x_t (see Line 9 in Algorithm B.1 in Supp.Materials). For sure, one needs to use at least standard binary tree techniques to make this search fast. The remaining algorithm operations whose computational complexity depends on d are those that create new boxes (see Lines 4, 19 and 20 in Algorithm B.1), they can be efficiently optimized as we describe below.

Second, the most memory consuming part of our algorithm is the storage of the state information for each box (i.e., the variables u, w, P, E,..). It is easy to see that we do not need to keep in memory the information for all the boxes. Indeed, one needs to store only information about the boxes that have had (in previous rounds) at least one context vector in them (since vectors from previous rounds have caused updates in values of u, w, P, and E). For the boxes that have had 0 context vectors inside, we can restore the values on u, w, P, E from parent boxes (due to Line 21 in Algorithm B.1 in Supp.Mat). Hence, in worst case, our algorithm requires $O(T)$ memory here.

Finally, the constants in the upper bound in Eq. (4) and some redundant actions can be optimized in a way similar to the one applied to the non-contextual setup in (Drutsa, 2017a).

6.5. Optimization of Regret with Other Loss

Let us change slightly our setup: assume the seller is not interested in minimization of the regret with pricing loss $(v(x_t) - a_t p_t$ in a round t), but in minimization of the cumulative symmetric loss $\sum_t |v(x_t) - p_t|$ as in (Mao et al., 2018, Sec.2.1); all other parts of the setup remains unchanged (i.e., *the buyer is strategic still*). In this case, our algorithm PELS can be used with effective upper bound on the symmetric regret. Namely, use PELS with parameters from Eq. (3), but take $g(m) \equiv 0$: exploitation rounds are not needed because we have no dramatic loss for an overguess. Since we have at least 1 round in a box, we get the worst-case number of boxes M s.t. $2^M \leq O(T^{1/d})$, by arguments similar to the ones in the proof of Th. 1. In order to obtain regret in each box of depth m bounded by $O(2^{-m})$, one needs to slightly modify the penalization rounds: offer the upper bound w^X instead of offering 1. In such round, the strategic buyer will not accept price as well (his value is lower than w^X), but the symmetric loss will be $O(|w^X - u^X|)$. Hence, *PELS modified in this way will have symmetric loss strategic regret of* $\sum_{m=0}^{M-1} 2^{md} O(2^{-m}) = O(T^{(d-1)/d})$, *which is tight for* $d > 1$ due to (Mao et al., 2018, Th.6).

7. Conclusions

We studied repeated contextual posted-price auctions with a strategic buyer that discounts his cumulative surplus and holds a private valuation in the form of a Lipschitz function of a d-dimensional context vector of a good. First, we closed the open research question on the existence of a no-regret pricing in this scenario by proposing a novel optimal learning algorithm that is horizon-independent and can act against the strategic buyer with tight regret upper bound of $\Theta(T^{d/(d+1)})$. Second, we generalized the value-localization approaches well know in the non-contextual setting to the multi-dimensional case. Finally, novel techniques were introduced: (a) the method to isolate penalization; and (b) the guarantee on the amount of lie from the buyer when he accepts an offered price.

References

Abernethy, J. D., Cummings, R., Kumar, B., Taggart, S., and Morgenstern, J. H. Learning auctions with robust incentive guarantees. In *Advances in Neural Information Processing Systems*, pp. 11587–11597, 2019.

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.

Agrawal, S., Daskalakis, C., Mirrokni, V., and Sivan, B. Robust repeated auctions under heterogeneous buyer behavior. *arXiv preprint arXiv:1803.00494*, 2018.

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.

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

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.

Baltaoglu, M. S., Tong, L., and Zhao, Q. Online learning of optimal bidding strategy in repeated multi-commodity auctions. In *Advances in Neural Information Processing Systems*, pp. 4507–4517, 2017.

Bikhchandani, S. Reputation in repeated second-price auctions. *Journal of Economic Theory*, 46(1):97–119, 1988.

Caillaud, B. and Mezzetti, C. Equilibrium reserve prices in sequential ascending auctions. *Journal of Economic Theory*, 117(1):78–95, 2004.

Carare, O. Reserve prices in repeated auctions. *Review of Industrial Organization*, 40(3):225–247, 2012.

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.

Cesa-Bianchi, N., Cesari, T., and Perchet, V. Dynamic pricing with finitely many unknown valuations. *arXiv preprint arXiv:1807.03288*, 2018.

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

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.

Deng, Y., Lahaie, S., and Mirrokni, V. A robust nonclairvoyant dynamic mechanism for contextual auctions. In *Advances in Neural Information Processing Systems*, pp. 8654–8664, 2019a.

Deng, Y., Schneider, J., and Sivan, B. Prior-free dynamic auctions with low regret buyers. In *Advances in Neural Information Processing Systems*, pp. 4804–4814, 2019b.

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

Drutsa, A. On consistency of optimal pricing algorithms in repeated posted-price auctions with strategic buyer. *CoRR*, abs/1707.05101, 2017a. URL http://arxiv.org/abs/1707.05101.

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

Drutsa, A. Weakly consistent optimal pricing algorithms in repeated posted-price auctions with strategic buyer. In *ICML'2018*, pp. 1318–1327, 2018.

Drutsa, A. Reserve pricing in repeated second-price auctions with strategic bidders. In *ICML'2020*, 2020.

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

Fu, H., Jordan, P., Mahdian, M., Nadav, U., Talgam-Cohen, I., and Vassilvitskii, S. Ad auctions with data. In *Algorithmic Game Theory*, pp. 168–179. Springer, 2012.

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

Golrezaei, N., Lin, M., Mirrokni, V., and Nazerzadeh, H. Boosted second-price auctions for heterogeneous bidders. 2017.

Golrezaei, N., Javanmard, A., and Mirrokni, V. Dynamic incentive-aware learning: Robust pricing in contextual auctions. In *Advances in Neural Information Processing Systems*, pp. 9756–9766, 2019.

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

Hart, O. D. and Tirole, J. Contract renegotiation and coasian dynamics. *The Review of Economic Studies*, 55(4):509–540, 1988.

Hartline, J. D. and Roughgarden, T. Simple versus optimal mechanisms. In *Proceedings of the 10th ACM conference on Electronic commerce*, pp. 225–234. ACM, 2009.

Hazan, E. and Megiddo, N. Online learning with prior knowledge. In *International Conference on Computational Learning Theory*, pp. 499–513. Springer, 2007.

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.

Huang, Z., Liu, J., and Wang, X. Learning optimal reserve price against non-myopic bidders. In *Advances in Neural Information Processing Systems*, pp. 2042–2052, 2018.

Hummel, P. Reserve prices in repeated auctions. *International Journal of Game Theory*, 47(1):273–299, 2018.

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.

Javanmard, A. and Nazerzadeh, H. Dynamic pricing in high-dimensions. *The Journal of Machine Learning Research*, 20(1):315–363, 2019.

Javanmard, A., Nazerzadeh, H., and Shao, S. Multi-product dynamic pricing in high-dimensions with heterogenous price sensitivity. *arXiv preprint arXiv:1901.01030*, 2019.

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.

Kleinberg, R., Slivkins, A., and Upfal, E. Multi-armed bandits in metric spaces. In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pp. 681–690. ACM, 2008.

Kleinberg, R. D. Nearly tight bounds for the continuum-armed bandit problem. In *Advances in Neural Information Processing Systems*, pp. 697–704, 2005.

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. and Schneider, J. Contextual search via intrinsic volumes. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 268–282. IEEE, 2018.

List, J. A. and Shogren, J. F. Price information and bidding behavior in repeated second-price auctions. *American Journal of Agricultural Economics*, 81(4):942–949, 1999.

Mao, J., Leme, R., and Schneider, J. Contextual pricing for lipschitz buyers. In *Advances in Neural Information Processing Systems*, pp. 5648–5656, 2018.

Mirrokni, V., Paes Leme, R., Tang, P., and Zuo, S. Non-clairvoyant dynamic mechanism design. 2017.

Mirrokni, V., Paes Leme, R., Tang, P., and Zuo, S. Optimal dynamic auctions are virtual welfare maximizers. *Available at SSRN*, 2018.

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.

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.

Peters, M. and Severinov, S. Internet auctions with many traders. *Journal of Economic Theory*, 130(1):220–245, 2006.

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.

Slivkins, A. Contextual bandits with similarity information. *The Journal of Machine Learning Research*, 15(1):2533–2568, 2014.

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 with different patience of the seller and the buyer. In *Advances in Neural Information Processing Systems*, pp. 939–951, 2019.

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.

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.

Zhiyanov, A. and Drutsa, A. Bisection-based pricing for repeated contextual auctions against strategic buyer. In *ICML'2020*, 2020.

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.