Bisection-Based Pricing for Repeated Contextual Auctions against Strategic Buyer

Anton Zhiyanov [1 2] **Alexey Drutsa** [1 2]

Abstract

We are interested in learning algorithms that optimize revenue in repeated contextual posted-price auctions where a single seller faces a single strategic buyer. In our setting, the buyer maximizes his expected cumulative discounted surplus, and his valuation of a good is assumed to be a fixed function of a d-dimensional context (feature) vector. We introduce a novel deterministic learning algorithm that is based on ideas of the Bisection method and has strategic regret upper bound of $O(\log^2 T)$. Unlike previous works, our algorithm does not require any assumption on the distribution of context information, and the regret guarantee holds for any realization of feature vectors (adversarial upper bound). To construct our algorithm we non-trivially adopted techniques of integral geometry to act against buyer strategicness and improved the penalization trick to work in contextual auctions.

1. Introduction

Revenue maximization is a permanent development goal in large Internet companies such as Real-time ad exchanges (RTB) and web search engines (Aggarwal et al., 2006; Vanunts & Drutsa, 2019). Ad inventory is usually sold by means of second-price auctions (He et al., 2013; Golrezaei et al., 2019; Drutsa, 2020a) and their variations for multiple ads per a web page, e.g., GSP (Varian, 2007) or VCG (Varian, 2009). Reserve prices are mainly used to extract optimal revenue in these online auction (Myerson, 1981; Cesa-Bianchi et al., 2013). In the case of a single advertiser (a frequent scenario in ad auctions (Amin et al., 2013; Mohri & Munoz, 2014; Drutsa, 2018)), a second-price auction with reserve prices reduces to a posted-price auction. In this mechanism, the auctioneer (the seller) sets a reserve price for an advertisement space (a good) and the advertiser (the buyer) decides whether to reject or accept this price: to bid below or above it (Kleinberg & Leighton, 2003).

In this work, we focus on a scenario when a single seller repeatedly interacts through posted-price auctions with the *same strategic* buyer (Amin et al., 2013; 2014). In each round, the currently offered good is described by a *context information* (feature vector), which is observed by both the buyer and the seller. The buyer holds a private (valuation) function to calculate his valuation of a good from its feature vector. For instance, this scenario models well an Ad exchange that sequentially offers different ad spaces to the same advertiser, who values differently visit of different users on different web sites. This dependence of the buyer's valuation on context is fixed and is unknown to the seller. We assume that the valuation function is from a certain parametric class and, for sake of exposition, this is the class of linear[1] models, following (Amin et al., 2014; Cohen et al., 2016; Leme & Schneider, 2018; Golrezaei et al., 2019).

The seller uses an *online learning algorithm* to set prices in each round based on previous decisions of the buyer and the observed context information of offered goods. This pricing algorithm is announced to the buyer in advance, and the strategic buyer seeks to maximize his expected cumulative discounted surplus (Amin et al., 2013; Drutsa, 2017b) with respect to some private distribution D of feature vectors (Amin et al., 2014; Golrezaei et al., 2019; Drutsa, 2020b). The seller is aimed to maximize her cumulative revenue over a finite time horizon T via regret minimization, i.e., she seeks for a pricing algorithm with a sublinear regret on T (a no-regret pricing) (Amin et al., 2013; 2014; Mohri & Munoz, 2014; Drutsa, 2018).

For this scenario, Amin et al. (2014) proposed the algorithm LEAP, which is based on Gradient Descent and has upper bound on expected strategic regret of the form $O(T^{2/3}\sqrt{\log T})$. Golrezaei et al. (2019) recently improved

[1]Yandex Research, Moscow, Russian Federation [2]Faculty of Mechanics and Mathematics, Moscow State University im. Lomonosova, Moscow, Russian Federation. Correspondence to: Anton Zhiyanov <zhiyanovap@gmail.com>, Alexey Drutsa <adrutsa@yandex.ru>.

Proceedings of the 37^{th} International Conference on Machine Learning, Vienna, Austria, PMLR 119, 2020.

[1]However, the results of our study hold for a variety of non-linear parametric functions (e.g., kernel models), see Sec. 7.

this upper bound to $O(\log^2 T)$ by introducing the algorithm CORP, which is based on techniques similar to EM-algorithm and requires some *assumptions on regularity* of the distribution D of feature vectors. Moreover, both LEAP's and CORP's upper bounds were provided for *expected* regret with respect to the distribution D.

In our study, we propose a novel pricing algorithm that exploits ideas of Bisection method (Binary search) and its extensions (Kleinberg & Leighton, 2003; Drutsa, 2017b; Leme & Schneider, 2018). This algorithm can be applied against the strategic buyer with regret upper bound of $O(\log^2 T)$ (see Theorem 1), which is currently known as the best asymptotic guarantee with respect to T, but, in contrast to the CORP algorithm, does not require any assumption on the buyer's distribution D. Moreover, our regret upper bound holds for *any realization of feature vectors* (worst-case guarantee with respect to context information) what is a stronger guarantee than the ones for expected regret in (Amin et al., 2014; Golrezaei et al., 2019). This result constitutes *the main contribution of our work*.

To construct our algorithm, we non-trivially upgrade the approaches from the fixed valuation (non-contextual) setup (Drutsa, 2017b) as follows: we inject penalizations not only after rejection of a price, but after an acceptance as well; dynamically increase the penalization rate; and introduce novel bounds on the buyer's contextual valuation (see Proposition 1 and 2). We also introduce novel techniques that are based on homothetic transformations and are vital to restrict lying ability of the buyer (see Lemmas 2 and 3). These methods are contributed by our work as well.

2. Setup of Repeated Contextual Auctions

We study the setting of *repeated contextual posted-price auctions* which is similar to the one described by Amin et al. (2014). 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 unit d-dimensional feature vector $x_t \in \mathbb{X} := \{x \in \mathbb{R}^d : \|x\| = 1\}$[2] also referred to as the *context* of the round t, $d \in \mathbb{N}$. The buyer holds a *private parameter vector* $\theta^* \in [0,1]^d$ used in the valuation function $f_{\theta^*} : \mathbb{X} \to \mathbb{R}_+$ to determine valuation of a good based on its context, where $f_{\theta^*}(x) = \langle x, \theta^* \rangle = \sum_{i=1}^d x_i \theta_i^* \; \forall x \in \mathbb{X}$. The valuation parameter is fixed over all rounds and is unknown to the seller. So, in each round t: both the seller and the buyer observe context information $x_t \in \mathbb{X}$; the seller offers a price $p_t \in \mathbb{R}_+$; and the buyer takes an allocation decision $a_t \in \{0,1\}$: $a_t = 1$, when he accepts to buy the currently offered good at price p_t, and $a_t = 0$, when he rejects it.

[2] From here on ℓ_2-norm is used: $\|x\| := \sqrt{\sum_{i=1}^d x_i^2}$ for $x \in \mathbb{R}^d$.

We consider the deterministic online learning scenario when the price p_t in a round $t \in \{1, \dots, T\}$ 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. The rule on how the prices $p_{1:T}$ are set in response to the buyer's decisions $a_{1:T}$ and observed features $x_{1:T}$ are referred to as a *pricing algorithm*.

Following, (Kleinberg & Leighton, 2003; Amin et al., 2013; 2014; Mohri & Munoz, 2014; Drutsa, 2017b; 2018), for a given play of the repeated game, the realized cumulative revenue $\sum_{t=1}^T a_t p_t$ of the seller is compared to the revenue that would have been earned by offering the buyer's *valuations* $\{v_t := \langle x_t, \theta^* \rangle\}_{t=1}^T$ if they were known in advance to the seller. This is made via the *regret* $\mathtt{Reg}(T, \mathcal{A}, \theta^*, a_{1:T}, x_{1:T}) := \sum_{t=1}^T (v_t - a_t p_t)$, where $\mathcal{A}$ is a used algorithm, $\theta^* \in [0,1]^d$ is a valuation parameter of the buyer that has made decisions $a_{1:T}$ for goods $x_{1:T}$.

We also assume that the seller's algorithm $\mathcal{A}$ is announced to the buyer *in advance*[4] (Amin et al., 2013; 2014; Mohri & Munoz, 2014; Drutsa, 2017b; Golrezaei et al., 2019). Hence, in each round t, our buyer acts strategically against this algorithm: given his belief about the distribution of feature vectors in future rounds, the buyer makes the optimal allocation decision[5] $a_t = a_t^{\mathrm{Opt}}(T, \mathcal{A}, \theta^*, \gamma, a_{1:t-1}, x_{1:t}, D)$, that maximizes his expected future γ-discounted *surplus* $\mathbb{E}_{x_{1:T} \sim D}[\sum_{s=t}^T \gamma^{s-1} a_s (v_s - p_s) \mid x_1, .., x_t]$, $\gamma \in [0,1]$ is the buyer's discount rate and D is a probability distribution over the feature domain $\mathbb{X}^T$ for goods $x_{1:T}$ (Amin et al., 2014).

Given T played rounds with realized feature vectors $x_{1:T}$, the *strategic regret* of the algorithm $\mathcal{A}$ that faced the strategic buyer with the valuation parameter $\theta^* \in [0,1]^d$ over T rounds is defined as

$$\mathtt{SReg}(T, \mathcal{A}, \theta^*, \gamma, x_{1:T}, D) := \\ = \mathtt{Reg}\big(T, \mathcal{A}, \theta^*, a_{1:T}^{\mathrm{Opt}}(T, \mathcal{A}, \theta^*, \gamma, x_{1:T}, D), x_{1:T}\big).$$

We are interested in pricing algorithms that have $o(T)$ strategic regret for the *worst-case* valuation parameter $\theta^* \in [0,1]^d$, the *worst-case* probability distribution D over the feature domain, and the *worst-case* realization of feature vectors $x_{1:T} \in \mathbb{X}^T$. Formally, an algorithm $\mathcal{A}$ is said to be *no-regret* when

$$\sup_{x_{1:T} \in \mathbb{X}^T, \theta^* \in [0,1]^d, D} \mathtt{SReg}(T, \mathcal{A}, \theta^*, \gamma, x_{1:T}, D) = o(T).$$

Note that SReg is a function of both a distribution D and context vectors $x_{1:T}$. The distribution dependency arises

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

[4] The seller is interested in commitment on algorithms, because non-commitment scenarios results in quite low revenue (Devanur et al., 2015; Vanunts & Drutsa, 2019; Golrezaei et al., 2019).

[5] Formally, the buyer applies an optimal strategy, where a strategy starting at a round t is a map of each possible observation of $p_{1:T}$ and goods $x_{1:T}$ to a sequence of allocation decisions $a_{t:T}$.

when the buyer maximizes his surplus and the context vectors dependency appear through the definition of the buyer's regret. We discuss this dependency in more details at Appendix C. Also we emphasize that an algorithm minimizing worst-case regret have to be deterministic (it only depends on the buyer's decisions and sequence of context vectors) and his regret does not depend on any randomness (it is not true for (Amin et al., 2014; Golrezaei et al., 2019) setup).

So, the optimization goal is to find an algorithm with the best possible asymptotic of the form $O(f(T))$, where $T \to \infty$.

3. Related Work and Background

Related work. There are two lines of works that are the most relevant to ours. The first one dealt with contextual pricing in repeated auctions (or multi-dimensional search in an online manner). 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 future surplus. The algorithms of (Cohen et al., 2016; Leme & Schneider, 2018; Javanmard & Nazerzadeh, 2019; Javanmard et al., 2019) search a valuation function as a parametric model, just as in our case. The second line of works studied our strategic setup with fixed private valuation, but in the non-contextual case (all goods are equal, $d = 1$): (Amin et al., 2013; Mohri & Munoz, 2014; Drutsa, 2017b; 2018; Schmidt, 1993; Hart & Tirole, 1988; Devanur et al., 2015; Immorlica et al., 2017; Vanunts & Drutsa, 2019; Drutsa, 2020a). The studies (Amin et al., 2014; Golrezaei et al., 2019; Drutsa, 2020b) lie at the intersection of both lines of works: their authors considered contextual repeated auctions where the seller interacts with the same strategic buyer. The algorithms of Amin et al. (2014) and Golrezaei et al. (2019) also explicitly assume that the valuation function is a particular parametric model, while the algorithm of Drutsa (2020b) learns the valuation function in a non-parametric way.

First, note that, in the setups considered by Amin et al. (2014); Golrezaei et al. (2019), *expected* regret $\sup_{\theta^* \in [0,1]^d, D} \mathbb{E}_{x_{1:T} \sim D} \texttt{SReg}(T, \mathcal{A}, \theta, \gamma, x_{1:T}, D)$ is minimized. This statement of the problem is not equivalent to ours. Recall that $\texttt{SReg}$ is a function of both a distribution D (the buyer maximizes his *surplus* with the fixed distribution D) and context vectors $x_{1:T}$. Then, in the case of *expected* regret, fixing the distribution, we compute $\mathbb{E}_{x_{1:T} \sim D} \texttt{SReg}(T, \mathcal{A}, \theta, \gamma, x_{1:T}, D)$ and lose an ability to variate vectors $x_{1:T}$. It is easy to see that an upper bound for our worst-case regret implies the same upper bound for expected regret (i.e., our regret guarantee is stronger). Second, in the algorithm CORP (Golrezaei et al., 2019), it is assumed that context vectors $x_{1:T} \sim D$ are independent and the probability distribution D is non-degenerate (i.e., the second moment matrix of the distribution D is positive definite). Our algorithm does not require these assumptions. Third, our algorithm is deterministic, while the ones from (Amin et al., 2014; Golrezaei et al., 2019) are not. Fourth, the non-parametric approach of Drutsa (2020b) can be applied to our scenario as well, but it will result in $O(T^{d/(d+1)})$ regret. In contrast, our algorithm provides better regret asymptotic: $O(\log^2 T)$, see Theorem 1.

Background on bisection-based pricing algorithms. Our scenario with $d = 1$ reduces to the setup of repeated *non*-contextual posted-price auctions earlier introduced in (Amin et al., 2013). In this case, pricing algorithms for worst-case regret minimization were well studied (Amin et al., 2013; Mohri & Munoz, 2014; Drutsa, 2017b; 2018). First, if the buyer cumulative utility is not discounted over rounds (i.e., the discount rate $\gamma = 1$), then the strategic regret is linear: its lower bound is $\Omega(T)$ (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 other discounts $\gamma \in [0, 1)$, 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 introduced for the non-contextual setup (Drutsa, 2017b; 2018). Their construction strongly relied on the technique of penalization (Mohri & Munoz, 2014; Drutsa, 2017b).

The recent work (Leme & Schneider, 2018) studied a setup of repeated contextual auctions with a buyer that holds a fixed linear valuation. But, their scenario considered the buyer that made decisions myopically (truthfully) in each round, what is only the special case of our setup with $\gamma = 0$. An optimal algorithm with tight truthful regret bound of $\Theta(\log\log T)$ were proposed in that work. It is based on a generalization of the non-contextual technique of reducing the size of a feasible search interval (Kleinberg & Leighton, 2003; Drutsa, 2017b) via the intrinsic volumes of the knowledge (search) set for the multidimensional parameter θ. Our algorithm also use this generalized search approach, but, in order to make it workable against the strategic buyer[6], we introduce novel multidimensional localization tools (otherwise, he may mislead the algorithm). There is a trade-off between the restriction of the buyer's lying ability and the increasing of the seller's regret. As a payment of this balance we incur regret upper bound of $O(\log^2 T)$, which is the best guarantee currently known for the contextual setup.

[6] Algorithms designed to act against a myopic buyer cannot be straightforwardly used against a strategic one: it was shown for the non-contextual setup (Drutsa, 2017b) and we discuss it in Sec. 5.

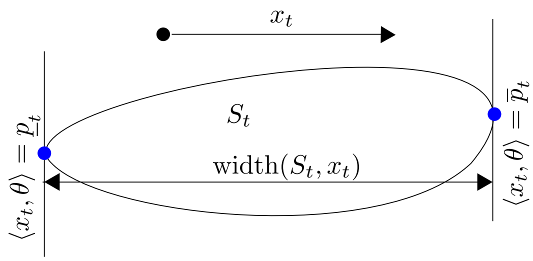

Figure 1. An illustration of notations for a knowledge set S_t: x_t is a current feature vector; $\langle x_t, \theta\rangle = \underline{p}_t$, $\langle x_t, \theta\rangle = \overline{p}_t$ are the lowest and highest hyperplanes orthogonal to x_t and intersecting S_t; $\mathrm{width}(S_t, x_t)$ is the width of S_t in the direction x_t.

4. Auxiliary definitions and statements

The algorithms we consider keep track of a feasible *knowledge set* aimed to locate the buyer's valuation parameter θ^*. This knowledge set is initialized by $[0,1]^d$ at the first round. Following (Cohen et al., 2016; Leme & Schneider, 2018), in each round t with context x_t, our algorithms update the current knowledge set by intersecting it with one of the halfspaces $\{\langle x_t, \theta\rangle \geq p\}$ or $\{\langle x_t, \theta\rangle \leq p\}$ for some parameter $p \in \mathbb{R}_+$. Namely, let S_t denote the current knowledge set, then the set S_{t+1} in the next round will remain the same or will be updated taking the form of either $S_t^-(p, x_t) := \{\theta \in S_t : \langle x_t, \theta\rangle \leq p\}$ or $S_t^+(p, x_t) := \{\theta \in S_t : \langle x_t, \theta\rangle \geq p\}$ depending on buyer allocation decision a_t. It is easy to see that $S_{t+1} \subseteq S_t$, and S_t is convex for any t (since S_1 is initially convex and further updates are intersections of convex sets).

Now we introduce several notations connected with geometric characteristic of a knowledge set. Given a context x_t in a round t, we set $\underline{p}_t := \min_{\theta\in S_t}\langle x_t, \theta\rangle$, and $\overline{p}_t := \max_{\theta\in S_t}\langle x_t, \theta\rangle$. For a set $S \subseteq [0,1]^d$, we define its diameter as $\mathrm{diam}(S) := \max_{\theta_1,\theta_2\in S}\|\theta_1 - \theta_2\|$ and its width in a direction x as $\mathrm{width}(S, x) := \max_{\theta\in S}\langle x, \theta\rangle - \min_{\theta\in S}\langle x, \theta\rangle$, where $x \in \mathbb{X}$ is a unit vector (for our approach, we have $\mathrm{width}(S_t, x_t) = \overline{p}_t - \underline{p}_t$). See Fig. 1 for an illustration of introduced notations.

Now we introduce a class of algorithms, for which we will provide Proposition 1 and 2 that are our localization techniques for θ^* and are vital to limit strategic behavior of the buyer.

Definition 1. A pricing algorithm is said to be *evaluating* if it keeps track of feasible sets $S_{1:T}$ s.t. $S_{t+1} \subseteq S_t$ for all rounds t, and it sets a price p_t s.t. $p_t \geq \underline{p}_t$ in each round t.

Finally, we present the notion of an intrinsic volume (Kleinberg, 1997) and introduce its most important properties for our algorithm. Let Conv_d be the class of compact convex sets in $\mathbb{R}^d$. For a compact convex set $K \in \mathrm{Conv}_d$, its j-th *intrinsic volume* is the coefficient $V_j(K)$, $j = 0, ..., d$, in the Steiner's formula for the (classical) volume of the (Minkowski) sum $K + \varepsilon B$: $\mathrm{Vol}(K + \varepsilon B) = \sum_{j=0}^{d} k_{d-j} V_j(K)\varepsilon^{d-j}$,where B is a unit ball, $\varepsilon > 0$, and k_{d-j} is the volume of the $(d-j)$-dimensional unit ball. The intrinsic volumes also can be interpreted as the expected volume of projection of the set K onto a random subspace of corresponding dimension (Leme & Schneider, 2018, Theorem 12). Note that $V_0(K) = 1$ and $V_d(K)$ equals to the classical volume of K, denoted by $\mathrm{Vol}(K)$; while V_{d-1} is an analogue of the perimeter of K for $d = 2$ and an analogue of the surface area K for $d = 3$. Intrinsic volumes have important properties (Leme & Schneider, 2018) similar to the ones of their analogues in the 2- and 3-dimensional cases. Namely, for any j, the map V_j is: (a) additive, i.e., $V_j(S_1 \cup S_2) = V_j(S_1) + V_j(S_2)\ \forall S_1, S_2 \in \mathrm{Conv}_d$ s.t. $S_1 \cap S_2 = \varnothing$; (b) monotone, i.e., $V_j(S_1) \leq V_j(S_2)\ \ \forall S_1, S_2 \in \mathrm{Conv}_d$ s.t. $S_1 \subseteq S_2$; and (c) j-homogenous, i.e., $V_j(\alpha S) = \alpha^j V_j(S)$ $\forall \alpha \in \mathbb{R}_+ \forall S \in \mathrm{Conv}_d$. The isoperimetric inequality holds: $(i!V_i(S))^{1/i} \geq ((i+1)!V_{i+1}(S))^{1/(1+i)}\ \forall S \in \mathrm{Conv}_d \forall i \geq 1$;

5. Localization of the buyer valuation

Drutsa (2017b, Theorem 4) showed that the strategic buyer may mislead algorithms that are designed to act effectively against myopic (truthful) buyers in non-contextual repeated auctions ($d = 0$ in our setup). Namely, if an algorithm uses a rejection of a currently offered price p as a signal that the buyer valuation is less than p (left consistency (Drutsa, 2017b)), then the strategic buyer can exploit this property to get surplus at least some $\epsilon > 0$ in each future round, what cause a linear strategic regret for the seller. Since the features are chosen adversarially in our contextual setting, the algorithms of (Cohen et al., 2016; Leme & Schneider, 2018) designed for a myopic (truthful) buyer[7] and used by the seller may have a linear regret against the strategic buyer. In order to reveal information on the buyer valuation from his binary decision a_t in a round t, the special trick (Drutsa, 2017b, Proposition 2) of penalization rounds is used in the non-contextual setting. We significantly expand this notion for our contextual case by using penalization not only in the case of rejection but also in the acceptance case:

Definition 2. For a pricing algorithm $\mathcal{A}$, a round τ is a *penalization* one, if its price $p_\tau = d + 1$ and, in the case of acceptance of this price, this algorithm will offer only the price $p_s = d + 1$ in all future rounds s. A round t is said to be the *start of r-length penalization*, if any next round s such that $t < s < t + r$ is a penalization one.

Unlike previous works (Drutsa, 2017b; Mohri & Munoz, 2014) we conduct penalization after any buyer decision.

[7] These algorithms use the logic of left consistency: if a price p_t is rejected for a context x_t, then $\langle x_t, \theta^*\rangle \leq p_t$.

Note that the strategic buyer will never accept the price in a penalization round since, otherwise, the price of any of future goods will be $d+1$, which exceeds his valuation. To get an upper bound on the buyer's valuation for a good x_t we prove contextual analogue of (Drutsa, 2017b, Prop. 2).

Proposition 1. *Let $\gamma \in [0,1)$, $\mathcal{A}$ be an evaluating pricing algorithm (see Def. 1) and a round t be the start of r-length penalization. Then, if the price p_t is rejected by the strategic buyer with a linear valuation function $\langle x, \theta^* \rangle$ and $\theta^* \in S_t$, then the following inequality on his current valuation holds:*

$$v_t := \langle x_t, \theta^* \rangle \le p_t + \frac{\gamma^r}{1-\gamma} \mathrm{diam}(S_t). \tag{1}$$

Proof. Let σ^{Opt} be the optimal strategy of the buyer with a start in the round t. Let σ' be the strategy, where the buyer accepts the good in the round t and rejects each future good. So, let $\mathtt{Sur}(\sigma)$ denote the buyer's expected future surplus (see Sec. 2) when he follows a strategy σ. Since σ^{Opt} is optimal, this implies that $\mathtt{Sur}(\sigma') \le \mathtt{Sur}(\sigma^{\mathrm{Opt}})$. By the definition, $\mathtt{Sur}(\sigma') = \gamma^{t-1}(\langle x_t, \theta^* \rangle - p_t)$, while the right-hand side of the previous inequality can be upper bounded:

$$\mathtt{Sur}(\sigma^{\mathrm{Opt}}) = \mathbb{E}[\sum_{s=t}^{T} \gamma^{s-1} a_s (v_s - p_s) \mid x_t, \sigma^{\mathrm{Opt}}] \le$$
$$\le \sum_{s=t+r}^{T} \gamma^{s-1} \mathrm{diam}(S_t) \le \frac{\gamma^{t+r-1}}{1-\gamma} \mathrm{diam}(S_t),$$

where we use that the rounds $t+1, \ldots, t+r-1$ will be penalization ones (the buyer will certainly reject them). Also we use the facts $\theta^* \in S_t$, $S_k \subseteq S_t$, and $p_k \ge \underline{p}_k$ for $k \ge t$ since our algorithm is evaluating. So, we get the following inequalities:

$$v_s - p_s \le \mathrm{width}(S_t, x_s) \le \mathrm{diam}(S_t), s = t+r, \ldots, T.$$

Combining all inequalities, we get $(v_t - p_t)\gamma^{t-1} \le \gamma^{t+r-1}\mathrm{diam}(S_t)/(1-\gamma)$, what implies Eq. 1 after dividing by γ^{t-1}. □

In the absence of context information for goods ($d = 1$), the previous bound was enough to restrict the buyer's lying ability (Drutsa, 2017b), since the knowledge set is unidimensional. In the multidimensional case, the situation is completely different, we cannot be sure in the implication: if a price p_t is accepted for a context, then $\langle x_t, \theta^* \rangle \ge p_t$, so we need to obtain a lower bound on the valuation.

Proposition 2. *Let $\gamma \in [0,1)$, $\mathcal{A}$ be an evaluating pricing algorithm (see Def. 1) and a round t be the start of a r-length penalization. Then, if the price p_t is accepted by strategic buyer with a linear valuation function $\langle x, \theta^* \rangle$ and $\theta^* \in S_t$, then the following inequality on his current valuation holds:*

$$v_t := \langle x_t, \theta^* \rangle \ge p_t - \frac{\gamma^r}{1-\gamma} \mathrm{diam}(S_t). \tag{2}$$

Proof. The proof of this proposition follows the logic of the previous one. Here we also consider the optimal strategy of the buyer σ^{Opt} starting on the round t. Since σ^{Opt} is optimal, we have $0 \le \mathtt{Sur}(\sigma^{\mathrm{Opt}})$, because we can consider the strategy where the buyer always reject the proposed price. We upper bound $\mathtt{Sur}(\sigma^{\mathrm{Opt}})$ as follows:

$$\mathtt{Sur}(\sigma^{\mathrm{Opt}}) = \mathbb{E}[\sum_{s=t}^{T} \gamma^{s-1} a_s (v_s - p_s) \mid x_t, \sigma^{\mathrm{Opt}}] \le$$
$$\le \gamma^{t-1}(v_t - p_t) + \frac{\gamma^{t+r-1}}{1-\gamma} \mathrm{diam}(S_t),$$

where we use that the rounds $t+1, \ldots, t+r-1$ will be penalization ones (the buyer will certainly reject them) and the buyer accepts the price p_t at the round t. Also we use the facts that $\theta^* \in S_t$ and knowledge sets $S_k \subseteq S_t$ and $p_k \ge \underline{p}_k$ for $k \ge t$ since our algorithm is evaluating: $v_s - p_s \le \mathrm{width}(S_t, x_s) \le \mathrm{diam}(S_t)$ for $s = t+r, ..., T$. Combining inequalities for $\mathtt{Sur}(\sigma^{\mathrm{Opt}})$, one can see $0 \le \gamma^{t-1}(v_t - p_t) + \gamma^{t+r-1}\mathrm{diam}(S_t)/(1-\gamma)$, what implies Eq. 2 after dividing by γ^{t-1}. □

6. Algorithm

In Algorithm 1 we present the pseudo-code of our algorithm. At first we note that our algorithm consists of two parts: the learning and exploiting phases. The exploiting phase occurs when $\mathrm{width}(S_t, x_t)/2 < 1/T$. The main idea of this phase as follows: after reducing of the knowledge set and well approximating of the valuation parameter θ^*, we propose the price p_t that will be definitely accepted by the buyer. Let us consider the learning phase (it occurs when $\mathrm{width}(S_t, x_t)/2 \ge 1/T$). Constants c_0 through c_{d-1} are defined so that $c_i = 1/(2^{i+1} i!), i = 0, ..., d-1$.

Following the isoperimetric inequality we keep track of the following "potentials" (they vary with t but we omit the subscript for notational convenience): $\varphi_i = \varphi_i(S_t) := (i! V_i(S_t))^{1/i}$. Since $S_1 = [0,1]^d$, their initial values are given by $\varphi_i(S_1) = (i!\binom{d}{i})^{1/i} < di < d^2$. Since those quantities are monotone non-increasing (w.r.t. t), they are always in the interval $[0, d^2)$. As in the bucketing procedure (Kleinberg & Leighton, 2003), we divide this interval into ranges of exponential decreasing length. The ranges will be $(l_{k+1}, l_k]$, where $l_k = d^2(1+1/d)^{-k}, k \in \mathbb{Z}_+$. For each potential φ_i, we keep track of the index $k(i)$ of the interval that contains the potential, i.e., $\varphi_i \in (l_{k(i)+1}, l_{k(i)}]$. By the isoperimetric inequality we know that: $\varphi_1 \ge \varphi_2 \ge \ldots \ge \varphi_d$ thus $k(1) \le k(2) \le \ldots \le k(d)$. So, each step of the learning phase of our algorithm corresponds to a

Algorithm 1 Pseudo-code of the algorithm

$C = \log_\gamma\left(\left(\frac{8(1+1/d)^d}{8(1+1/d)^d-1}\right)^{1/d} - 1\right) + \log_\gamma(1-(3/4)^{1/d})$
$t = 1; \quad S_1 = [0,1]^d$
while $t \le T$ **do**
  the seller receives a unit context vector $x_t \in \mathbb{R}^d$
  $w := \mathrm{width}(S_t, x_t)/2$
  if $w < 1/T$ **then**
    offer price $p_t := \underline{p}_t$
    $S_{t+1} := S_t; \quad t := t+1$
  else
    for all $i \in \{1, \dots, d\}$ **do**
      $\varphi_i := (i! V_i(S_t))^{1/i}$
      find $k(i)$ s.t. $\varphi_i \in (l_{k(i)+1}, l_{k(i)}]$
      if $V_i(S_t) - V_i(S_t^+(\overline{p}_t, x_t)) > l^i_{k(i)+1}/(4i!)$ **then**
        find p_i s.t. $V_i(S_t) - V_i(S_t^+(p_i, x_t)) = l^i_{k(i)+1}/(4i!)$
      else
        $p_i := \overline{p}_t$
      end if
      $K_i := \{\theta \in S_t : \langle x_t, \theta\rangle = p_i\}$
      $L_i := (V_i(K_i)/c_i)^{1/i}$ (define $L_0 := \infty$)
      $M(i) := \max\{j : k(i) = k(j)\}$
    end for
    find j s.t. $L_{j-1} > w \ge L_{M(j)}$
    $J := M(j)$
    $m_t := \lceil \log_\gamma(1-\gamma) + \log_\gamma w - \log_\gamma \mathrm{diam}(S_t) + C \rceil$
    offer price $p_t := \max(\underline{p}_t, p_J - \gamma^{m_t}\mathrm{diam}(S_t)/(1-\gamma))$
    make penalization for m_t rounds
    if price p_t accepted at the round t **then**
      $S_{t+m_t} := S_t^+(p_t - \gamma^{m_t}\mathrm{diam}(S_t)/(1-\gamma), x_t)$
    else
      $S_{t+m_t} := S_t^-(p_t + \gamma^{m_t}\mathrm{diam}(S_t)/(1-\gamma), x_t)$
    end if
    $t := t + m_t$
  end if
end while

state, described by a series of the numbers $\{k(i)\}_{i=1}^d$. Similar to (Cohen et al., 2016; Leme & Schneider, 2018) we try to reduce the knowledge set. Note that we use penalization to evaluate the difference between v_t and p_t (and shrink the knowledge set after that optimally). More penalization improves this evaluation, but increases the regret. This approach allows us to shrink the gap between $S_t^+(p_J, x_t)$ and $S_t^+(p_t - \gamma^{m_t}\mathrm{diam}(S_t)/(1-\gamma), x_t)$, increasing $V_J(S_t) - V_J(S_t^+(p_t - \gamma^{m_t}\mathrm{diam}(S_t)/(1-\gamma), x_t))$. It is important to us, since the last difference allows us to bound the number of steps required to leave the current range in the acceptance case. Thus, in Sec. 6.2, we select m_t so that the number of such steps and the regret are small.

6.1. Properties of the Algorithm

The structure of our algorithm resembles the algorithm (Leme & Schneider, 2018, Algorithm 5), what causes the presence of similar properties. Formally, the guarantees of (Leme & Schneider, 2018) cannot be directly used in our case, and, hence, the following statements need to be proven for our algorithm separately. However, the proofs of these properties are very similar to those that are present in (Leme & Schneider, 2018). Therefore, we briefly overview them, while the full proofs can be found in Appendix A.1.

Statement 1. *It is always possible to choose p_i s.t. $V_i(S_t) - V_i(S_t^+(p_i, x_t)) = l^i_{k(i)+1}/(4i!)$, if $V_i(S_t) - V_i(S_t^+(\overline{p}_t, x_t)) > l^i_{k(i)+1}/(4i!)$. Also there is an index j s.t. $L_{j-1} > w \ge L_{M(j)}$.*

The following statement will help[8] us derive that, if we receive a rejection, the quantity φ_J jumps from the range $(l_{k(J)+1}, l_{k(J)}]$ to the next range $(l_{k(J)+2}, l_{k(J)+1}]$.

Statement 2. *For the price p_J chosen in Algorithm 1, the following inequalities hold: $[J! V_J(S_t^-(p_J, x_t))]^{1/J} \le l_{k(J)+1}$ and $V_J(K_J) \le l^J_{k(J)+1}/(2J!)$.*

Statement 3 allows us to bound the number of times we can potentially take acceptance before $k(J)$ changes; bounds the width of the knowledge set; and gives us an asymptotic upper bound on the number of learning steps of the algorithm.

Statement 3. *For the index J and the price p_J chosen in Algorithm 1, the following equation and inequalities hold: (a) $V_J(S_t) - V_J(S_t^+(p_J, x_t)) = l^J_{k(J)+1}/(4J!)$; (b) $w \le 2l_{k(J)}$; and (c) $k(J) \le 4(d\log_2 dT + 1)$ in the rounds where $w \ge 1/T$.*

6.2. Localization technique in the algorithm

This is where the common features of our algorithm Algorithm 1 and the one of (Leme & Schneider, 2018) end. We noted in Section 5 that the strategic buyer may mislead algorithms that are designed to act effectively against truthful buyers. As one can see from our algorithm the main technique that allows us to fight with a strategic buyer is penalization rounds. First of all we show that our algorithm is evaluating.

Lemma 1. *Algorithm 1 is evaluating and $\theta^* \in S_t$ for all t.*

Proof. From the construction of the algorithm, $p_t \ge \underline{p}_t$ and $S_{t+1} \subseteq S_t$ for all rounds t. Now we proof by induction that $\theta^* \in S_t$ for all t. In the first round, it follows from the definition of S_1. Let $\theta^* \in S_t$ at a round t. If this round is exploiting, then $S_{t+1} = S_t$ and $\theta^* \in S_{t+1}$. Otherwise, if the round t is learning, then we propose the price p_t to the buyer and make penalization for m_t rounds. These rounds give us bounds (see Proposition 1, 2) between the real and the offered price for the good x_t:

[8] As one can see later in Corollary 1, the price proposed to the buyer is $p_t = p_J - \gamma^{m_t}\mathrm{diam}(S_t)/(1-\gamma) \ge \underline{p}_t$ and, thus, after rejection we take $S_{t+m_t} = S_t^-(p_t + \gamma^{m_t}\mathrm{diam}(S_t)/(1-\gamma) = p_J, x_t)$.

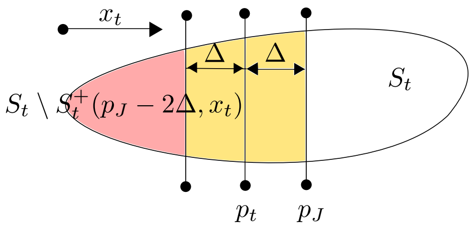

Figure 2. Illustration of the set $S_t \setminus S_t^+(p_J - 2\Delta, x_t)$ (the J-th intrinsic volume of this set influences on change of potential φ_J after acceptance). Here we use the following notation: $\Delta := \gamma^{m_t}/(1-\gamma)\,\mathrm{diam}(S_t)$.

(a) in the case of rejection and following m_t penalization rounds, we have: $v_t - p_t \leq \gamma^{m_t}\mathrm{diam}(S_t)/(1-\gamma)$ and $\theta^* \in S_t^-(p_t + \gamma^{m_t}\mathrm{diam}(S_t)/(1-\gamma), x_t) = S_{t+m_t}$; (b) in the case of acceptance and following m_t penalization rounds, we have: $v_t - p_t \geq -\gamma^{m_t}\mathrm{diam}(S_t)/(1-\gamma)$ and $\theta^* \in S_t^+(p_t - \gamma^{m_t}\mathrm{diam}(S_t)/(1-\gamma), x_t) = S_{t+m_t}$. □

As we will show later in Corollary 1, the proposed prise is $p_t = p_J - \gamma^{m_t}\mathrm{diam}(S_t)/(1-\gamma) \geq \underline{p}_t$. In the case of a rejection round t, this will allow us to make the conclusion that $\theta^* \in S_t^-(p_J, x_t)$ and φ_J will jump from the range $(l_{k(J)+1}, l_{k(J)}]$ to the next range $(l_{k(J)+2}, l_{k(J)+1}]$ (see Statement 2). On the other hand, after an acceptance round, see Fig. 2, we will know that $\theta^* \in S_t^+(p_J - 2\gamma^{m_t}\mathrm{diam}(S_t)/(1-\gamma), x_t)$, but we can only bound the difference $V_J(S_t) - V_J(S_t^+(p_J, x_t))$, from Statement 3. So, one needs one more result to be able to process acceptance rounds. Namely, we have to bound the gap between $S_t^+(p_J, x_t)$ and $S_t^+(p_J - 2\gamma^{m_t}\mathrm{diam}(S_t)/(1-\gamma), x_t)$ in order to have a tool that bounds the number of times we can potentially face acceptance before $k(J)$ changes.

We bound the distance between the top of S_t and the hyperplane containing K_J. Here we use an idea: since $V_J(S^+(p_J, x_t))$ has a lower bound (it follows from the choice of p_J and from the definition of $k(J)$, see Eq. 4), this distance cannot be arbitrary small. We abbreviate $S_t^-(p_J, x_t)$ and $S_t^+(p_J, x_t)$ by S_t^- and S_t^+ respectively.

Lemma 2. *Let P_1 and P_2 be hyperplanes orthogonal to the vector x_t: $P_1 := \{\theta \in \mathbb{R}^d : \langle x_t, \theta\rangle = p_J\}$ and $P_2 := \{\theta \in \mathbb{R}^d : \langle x_t, \theta\rangle = \overline{p}_t\}$. Then, for the distance l between P_1 and P_2, the following inequality holds: $l \geq 2w[1-(3/4)^{1/J}]$.*

Proof. See Fig. 3. At first, from Statement 3, we get

$$V_J(S_t) - V_J(S_t^+) = l_{k(J)+1}^J/(4J!). \tag{3}$$

Since $\varphi_J(S_t) \in (l_{k(J)+1}, l_{k(J)}]$, we have $V_J(S_t) \geq l_{k(J)+1}^J/J!$. Substituting this inequality in the Eq. 3 we get

$$V_J(S_t^+) \geq 3l_{k(J)+1}^J/(4J!). \tag{4}$$

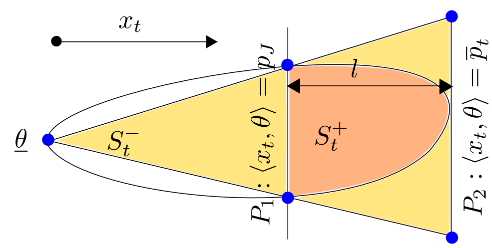

Figure 3. Here one can see the distance l between the hyperplane P_2 and the hyperplane P_1, considering K_J. Since the set S_t is convex, S_t^+ is a subset of a cone with a top $\underline{\theta}$ and a bottom P_2.

Now we give an upper bound on $V_J(S_t^+)$. Let $\underline{\theta}$ be a point from $\mathrm{argmin}\{\langle x_t, \theta\rangle : \theta \in S_t\}$ (in particular, $\langle x_t, \underline{\theta}\rangle = \underline{p}_t$) and $\mathrm{Cone} := \mathrm{Cone}(K_J, \underline{\theta})$ be the cone whose base is $K_J \subseteq P_1$ and whose top is $\underline{\theta}$. Since S_t is convex, we have that $\mathrm{Cone} \subseteq S_t^-$. Let $\mathrm{Hom} : \mathbb{R}^d \to \mathbb{R}^d$ be the homothetic transformation with center at the point $\underline{\theta}$ and scale coefficient $2w/(2w-l)$. For $\mathrm{Hom}(\mathrm{Cone})$, convexity of S_t implies that $S_t^+ \subseteq (\mathrm{Hom}(\mathrm{Cone}) \setminus \mathrm{Cone}) \cup K_J$ and, thus,

$$V_J(S_t^+) \leq V_J(\mathrm{Hom}(\mathrm{Cone})) - V_J(\mathrm{Cone}) + V_J(K_J). \tag{5}$$

The last inequality holds, since the intrinsic volumes are additive and monotone. Also from homogeneity of the intrinsic volumes we have $V_J(\mathrm{Hom}(\mathrm{Cone})) = (2w/(2w-l))^J V_J(\mathrm{Cone})$. Substituting it in the Eq. 5 and using the monotone we get

$$V_J(S_t^+) \leq V_J(\mathrm{Cone})[(2w/(2w-l))^J - 1] + V_J(K_J) \leq$$
$$\leq V_J(S_t^-)[(2w/(2w-l))^J - 1] + V_J(K_J),$$

here we used that $\mathrm{Cone} \subseteq S_t^-$, $V_J(\mathrm{Cone}) \leq V_J(S_t^-)$.

It remains to get an upper bound on $V_J(S_t^-)$ and $V_J(K_J)$ to bound the expression for $V_J(S_t^+)$. The upper bound on $V_J(K_J)$ follows from Statement 2. Since V_J is additive and $S_t^- = (S_t \setminus S_t^+) \cup K_J$, it follows that $V_J(S_t^-) = V_J(S_t) - V_J(S_t^+) + V_J(K_J) \leq l_{k(J)+1}^J/(4J!) + l_{k(J)+1}^J/(2J!)$. Here we used Eq. 3 to bound $V_J(S_t) - V_J(S_t^+)$. In result, we have

$$V_J(S_t^+) \leq \frac{3l_{k(J)+1}^J}{4J!}\left(\frac{2w}{2w-l}\right)^J - \frac{l_{k(J)+1}^J}{4J!}. \tag{6}$$

Combining Eq. 4 and Eq. 6, we get our inequality. □

Now we show that $p_J - \gamma^{m_t}\mathrm{diam}(S_t)/(1-\gamma) \geq \underline{p}_t$. In order to do this, we will prove a more general result for an arbitrary value h (instead of $\gamma^{m_t}\mathrm{diam}(S_t)/(1-\gamma)$):

Lemma 3. *Let $P_{3,h} := \{\theta \in \mathbb{R}^d : \langle x_t, \theta\rangle = p_J - h\}$. If the inequality*

$$h \leq 2w\left[\left(\frac{4(1+1/d)^J}{4(1+1/d)^J - 1}\right)^{1/J} - 1\right]\left[1-(3/4)^{1/J}\right]$$

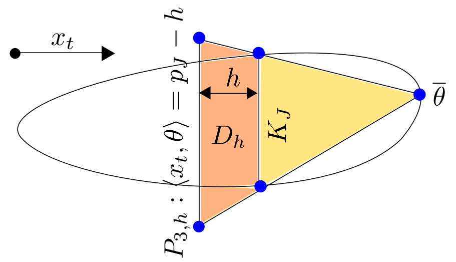

Figure 4. Here one can see the distance h between the hyperplane $P_{3,h}$ and K_J. This cone is a result of homothety of a cone with the same top and a bottom P_1. D_h is a subset of a cone with a top $\overline{\theta}$ and a bottom $P_{3,h}$. This cone is a result of homothety of the cone with the same top and a bottom K_J.

holds then $P_{3,h} \cap S_t \neq \varnothing$.

Proof. See Fig. 4. Let $\overline{\theta}$ be a point from $\operatorname{argmax}\{\langle x_t, \theta\rangle : \theta \in S_t\}$ (in particular, $\langle x_t, \overline{\theta}\rangle = \overline{p}_t$) and $\mathrm{Cone} := \mathrm{Cone}(K_J, \overline{\theta})$ be the cone whose base is K_J and whose top is $\overline{\theta}$. Let $\mathrm{Hom}_h : \mathbb{R}^d \to \mathbb{R}^d$ be the homothetic transformation with center at the point $\overline{\theta}$ and scale coefficient $(l+h)/l$, where l is defined in Lemma 2. Define $D_h := \mathrm{Hom}_h(\mathrm{Cone}) \cap \{\theta \in \mathbb{R}^d : p_J - h \le \langle x_t, \theta\rangle \le p_J\}$. Note that, since S_t is convex and K_J is the base of Cone, it is enough to show that $V_J(D_h) \le V_J(S_t^-)$ to prove our statement. From homogeneity of the intrinsic volumes we have $V_J(\mathrm{Hom}_h(\mathrm{Cone})) = ((l+h)/l)^J V_J(\mathrm{Cone})$. Using $\mathrm{Cone} \subseteq S_t^+$ and $D_h = (\mathrm{Hom}_h(\mathrm{Cone}) \setminus \mathrm{Cone}) \cup K_J$, we get

$$V_J(D_h) \le V_J(\mathrm{Cone})[((l+h)/l)^J - 1] + V_J(K_J) \le \\ \le V_J(S_t^+)[((l+h)/l)^J - 1] + V_J(K_J),$$

here we also used that the intrinsic volumes are monotone and additive. So, we just have to show that

$$V_J(K_J) + V_J(S_t^+)[\left(\frac{l+h}{l}\right)^J - 1] \le V_J(S_t^-) \quad (7)$$

The upper bound on $V_J(S_t^+)$ follows from Eq. 3

$$V_J(S_t^+) = V_J(S_t) - \frac{l_{k(J)+1}^J}{4J!} \le \frac{l_{k(J)}^J}{J!} - \frac{l_{k(J)+1}^J}{4J!}, \quad (8)$$

since $\varphi_J(S_t) \in (l_{k(J)+1}, l_{k(J)}]$. Substituting Eq. 8 in Eq. 7 and using Eq. 3 after moving the term $V_J(K_J)$ in the right side of the inequality: $V_J(S_t^-) - V_J(K_J) = V_J(S_t) - V_J(S_t^+) = l_{k(J)+1}^J/(4J!)$, we have to prove that $h \le l[\left(4l_{k(J)}^J / \left(4l_{k(J)}^J - l_{k(J)+1}^J\right)\right)^{1/J} - 1]$. It remains to substitute the bound of l from Lemma 2 and the definition of l_k. ☐

Corollary 1. *Let the number of penalization rounds* m_t *be given as in Alg. 1. Then* $p_J - \gamma^{m_t}\mathrm{diam}(S_t)/(1-\gamma) \ge \underline{p}_t$.

Proof Sketch. The full prove is in Appendix A.2.1. There we apply Lemma 3, where $h = \gamma^{m_t}\mathrm{diam}(S_t)/(1-\gamma)$. ☐

Now we are able to bound the difference between S_t and $S_t^+(p_t - \gamma^{m_t}\mathrm{diam}(S_t)/(1-\gamma), x_t)$, where $p_t = p_J - \gamma^{m_t}\mathrm{diam}(S_t)/(1-\gamma)$ is the price proposed to the buyer in the algorithm, since $p_J - \gamma^{m_t}\mathrm{diam}(S_t)/(1-\gamma) \ge \underline{p}_t$ from Corollary 1. This difference will help us to bound the number of steps that we make before leaving the current bucket in the acceptance case.

Corollary 2. *Let the number of penalization rounds* m_t *be given as in Algorithm 1. Then*

$$V_J(S_t) - V_J(S_t^+(p_J - 2\frac{\gamma^{m_t}}{1-\gamma}\mathrm{diam}(S_t), x_t)) \ge \frac{l_{k(J)+1}^J}{8J!}.$$

Proof Sketch. The full proof is in Appendix A.2.2. We use the notations of Lemma 3 and the abbreviation S_t^+ for $S_t^+(p_J, x_t)$. Let $h = 2\gamma^{m_t}/(1-\gamma)\mathrm{diam}(S_t)$. To prove this statement we have to get an upper bound for $S_t^+(p_J - h, x_t)$. Indeed, since S_t is convex, we have $S_t^+(p_J - h, x_t) \subseteq S_t^+ \cup D_h \subseteq S_t^+ \cup (\mathrm{Hom}_h(\mathrm{Cone}) \setminus \mathrm{Cone})$. Since $\mathrm{Cone} \subseteq S_t^+$, and using monotone and homogeneity of the intrinsic volumes, we get $V_J(\mathrm{Hom}_h(\mathrm{Cone})) - V_J(\mathrm{Cone}) = V_J(\mathrm{Cone})([(l+h)/l]^J - 1) \le V_J(S_t^+)([(l+h)/l]^J - 1)$.

Thus, using Eq. 3 to bound $V_J(S_t) - V_J(S_t^+)$, we get $V_J(S_t) - V_J(S_t^+(p_J - h, x_t)) \ge V_J(S_t) - V_J(S_t^+) - V_J(S_t^+)[((l+h)/l)^J - 1] = l_{k(J)+1}^J/(4J!) - V_J(S_t^+)[((l+h)/l)^J - 1]$. Therefore, we have to check that

$$V_J(S_t^+(p_J, x_t)) \left[\left(\frac{l+h}{l}\right)^J - 1\right] \le \frac{l_{k(J)+1}^J}{4J!} - \frac{l_{k(J)+1}^J}{8J!},$$

that follows from the definition of m_t and Eq. 8. ☐

Remark 1. Let the number of penalization rounds m_t be given as in Alg. 1. Then $m_t \le O(\log_\gamma(1-\gamma) + \log_\gamma dT)$[9].

Proof Sketch. It follows from the fact that penalization rounds occur in the case $w \ge 1/T$ and the inequality $\mathrm{diam}(S_t) \le d+1$. See App. A.2.3 for more details. ☐

6.3. Main result

Theorem 1. *Let* $\mathcal{A}$ *be Algorithm1 for a fixed time horizon* T. *Then the seller' total regret* $\mathtt{SReg}(T, \mathcal{A}, \theta^*, \gamma, x_{1:T}, D)$ *has an upper bound of*

$$O(d^3 \log_2^2(dT)\log_2(\gamma) + d^3 \log_2(dT)\log_\gamma(1-\gamma)) \quad (9)$$

for all $\theta^* \in [0,1]^d, x_{1:T} \in \mathbb{X}^T$ *and all distributions* D *over the feature domain* $\mathbb{X}^T$.

[9]Hereafter, we leave γ, T and d dependency in asymptotic.

Proof. We sum regret in different cases. The first one, when $w < 1/T$ and the algorithm sets the price $\underline{p}_t$. In this case, the algorithm always sells since $\underline{p}_t \le v_t$ and there is no need to lie to the buyer in such rounds, because it does not affects on the next rounds. The regret is at most $2w \le 2/T$ per round, so the total regret is at most 2 in this case.

The second case comes when $w \ge 1/T$, where we propose the price $p_t = p_J - \gamma^{m_t}\mathrm{diam}(S_t)/(1-\gamma)$ (see Corollary 1) and the algorithm receives a rejection and does not sell. We have m_t penalization rounds after such round. Then θ^* belongs to $S_{t+m_t} = S_t^-(p_t + \gamma^{m_t}\mathrm{diam}(S_t)/(1-\gamma) = p_J, x_t)$ (see Lemma 1), φ_J goes from the bucket $(l_{k(J)+1}, l_{k(J)}]$ to the next bucket $(l_{k(J)+2}, l_{k(J)+1}]$ (it follows from Statement 2). Since $m_t = O(\log_\gamma(1-\gamma) + \log_\gamma(dT))$ by Remark 1, we get regret at most $(d+1)(1+m_t) = O(d\log_\gamma(1-\gamma) + d\log_\gamma(dT))$ (the loss of rejection and penalization rounds). Since $k(J) \le O(d \log_2 dT)$ (see Statement 3), this can happen at most $O(d\log_2 dT)$ times for each index J. Since there are d such indices, the total regret of this case has an upper bound of $O(d^3 \log_2(dT)\log_\gamma(1-\gamma) + d^2\log_2^2(dT)\log_2(\gamma))$.

The final case is when the algorithm receives an acceptance. The loss in this case is bounded by $2w$ in the round of acceptance and by $d+1$ in the penalization rounds. Let us fix the selected index J and the index $k(J)$. The loss after the acceptance and m_t penalization rounds is at most $2w + m_t(d+1) \le 4l_{k(J)} + m_t(d+1)$ by Statement 3. When it happens, $\theta^* \in S_{t+m_t} = S_t^+(p_t - \gamma^{m_t}\mathrm{diam}(S_t)/(1-\gamma), x_t)$ by Lemma 1. Then, using Corollary 2, the J-th intrinsic volume decreases by at least $l^J_{k(J)+1}/(8J!)$. Therefore, the total number of times it (acceptance and following penalization) happens before leaving the current range is $l^J_{k(J)}/(l^J_{k(J)+1}/8) = 8(1+1/d)^J = O(1)$. So, the total regret of such event is at most $(2w + m_t(d+1))l^J_{k(J)}/(l^J_{k(J)+1}/8) \le (4l_{k(J)} + m_t(d+1))\,8(1+1/d)^J = O(d\log_\gamma(1-\gamma) + d\log_2(dT))$. By summing over all d possible values of J and all $O(d\log_2 dT)$ possible values of $k(J)$ we obtain the total loss of $O(d^3\log_2(dT)\log_\gamma(1-\gamma) + d^2\log_2^2(dT)\log_2(\gamma))$. ☐

7. Discussion

Asymptotic: Note that the better a good should be described the higher the dimension of a feature vector x_t is required. So, it is important that our algorithm has polynomial regret on the dimension parameter d. Also note that this regret blows up as γ tends to 1, but this is an expected behavior since: (a) in non-contextual cases, upper bounds have the same behavior (Mohri & Medina, 2014; Drutsa, 2017b; 2018); (b) for $\gamma = 1$, there does not exist a no-regret algorithm (Amin et al., 2013).

Squaring trick: One can see that our algorithm requires knowledge of the horizon T. To be free of this assumption, we apply the standard technique "squaring trick" (Lin et al., 2015; Cohen et al., 2016): if the learning algorithm has regret $O(\log^c T)$, $c > 0$ then running of independent instances of this algorithm during subsequent increasing phases (i-th phase has length 2^{2^i}) will have regret $O(\log^c T)$ as well. This trick is applicable for the strategic buyer, since his decisions during current phase does not affect on other (Drutsa, 2017b). We formally discuss it in Appendix B.1.

Extension to nonlinear models: Our setup focuses on linear valuation model, but it is easy to generalize our analysis to some of nonlinear models. Let us consider valuation model $v_t := \phi(\langle\psi(x_t), \theta^*\rangle)$, where $\psi : \mathbb{X} \to \mathbb{X}$ is a mapping and $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ is an increasing function s.t. $\phi(x) + h \le \phi(x+h)$ for all $x, h \ge 0$. After the change of variables $\tilde{x}_t := \psi(x_t)$ and $\tilde{v}_t := \phi^{-1}(v_t)$, we return to the problem discussed, but with $\tilde{v}_t = \langle\tilde{x}_t, \theta^*\rangle$. We discuss it more formally in Appendix B.2.

Independence of γ: Let us assume that we do not know the true value of the parameter γ, but its upper bound γ_0 (s.t. $\gamma \le \gamma_0 < 1$) is known. Note that bounds from Proposition 1, 2 are monotonous in the parameter γ and blows up when $\gamma \to 1$. Therefore, the bounds from these statements are true for γ_0. Using Algorithm 1 for the parameter γ_0, we get that the asymptotic of the seller's regret still has the form of $O(\log^2 T)$. From this, we can conclude that our algorithm is also applicable in the situation when the true value of the parameter γ is unknown and we just have its upper bound. We discuss it more formally in Appendix B.3.

Possible ways to optimize constants in the regret upper bound Eq. (9) can be seen in (Drutsa, 2017a), where such constants have been minimized for the non-contextual algorithm RPPFES.

8. Conclusion

We studied repeated contextual posted-price auctions with a strategic buyer that discounted his cumulative surplus and held a private valuation in the form of a parametric function of a d-dimensional context vector of a good. First, we proposed a novel learning algorithm that can act against the strategic buyer and has the best current upper bound of $O(\log^2 T)$ for the seller's regret. This bound is similar to the bound of the algorithm CORP, but our algorithm is deterministic and works for all distributions D over $\mathbb{X}^T$ (unlike the algorithm CORP). Second, we generalized the value-localization approaches well know in the non-contextual setting to the multidimensional case. Finally, novel techniques were introduced: (a) the application of penalization rounds in the case of acceptance; (b) dynamic increase of the penalization rate; and (c) homothetic transformation analysis to restrict lying ability of the buyer.

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, A., and Motwani, R. Truthful auctions for pricing search keywords. In *Proceedings of the 7th ACM conference on Electronic commerce*, pp. 1–7. ACM, 2006.

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.

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 non-clairvoyant 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*, 2020a.

Drutsa, A. Optimal non-parametric learning in repeated contextual auctions with strategic buyer. In *ICML'2020*, 2020b.

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.

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.

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. *Introduction to geometric probability*. Cambridge University press, 1997.

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. 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.

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

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.

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.

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.