Reserve Pricing in Repeated Second-Price Auctions with Strategic Bidders

Alexey Drutsa [1 2]

Abstract

We study revenue optimization learning algorithms for repeated second-price auctions with reserve where a seller interacts with multiple strategic bidders each of which holds a fixed private valuation for a good and seeks to maximize his expected future cumulative discounted surplus. We propose a novel algorithm that has strategic regret upper bound of $O(\log\log T)$ for worst-case valuations. This pricing is based on our novel transformation that upgrades an algorithm designed for the setup with a single buyer to the multi-buyer case. We provide theoretical guarantees on the ability of a transformed algorithm to learn the valuation of a strategic buyer, which has uncertainty about the future due to the presence of rivals.

1. Introduction

Revenue maximization is one of fundamental development directions in major Internet companies that have their own online advertising platforms (Gomes & Mirrokni, 2014; Balseiro et al., 2015; Agarwal et al., 2014; Drutsa, 2017b; Hummel, 2018). Most part of ad inventory is sold via widely applicable second price auctions (He et al., 2013; Mohri & Medina, 2014) and their generalizations like GSP (Varian, 2007; 2009; Varian & Harris, 2014; Sun et al., 2014). Adjustment of reserve prices plays a central role in revenue optimization here: their proper setting is studied both by game-theoretical methods (Myerson, 1981; Agrawal et al., 2018) and by machine learning approaches (Nisan et al., 2007; Cesa-Bianchi et al., 2013; Mohri & Medina, 2014; Paes Leme et al., 2016).

In our work, we focus on a scenario where the seller *repeatedly* interacts through a second-price auction with M strategic bidders (referred to as buyers as well). Each buyer participates in each round of this game, holds a *fixed* private valuation for a good (e.g., an ad space), and seeks to maximize his expected future discounted surplus given his beliefs about the behaviors of other bidders. The seller applies a deterministic online learning algorithm, which is announced to the buyers in advance and, in each round, selects individual reserve prices based on the previous bids of the buyers. The seller's goal is to maximize her revenue over a finite horizon T through *regret* minimization for *worst-case* valuations of the bidders (Mohri & Munoz, 2014; Drutsa, 2018). Thus, the seller seeks for a *no-regret* pricing algorithm.

To the best of our knowledge, no existing study investigated *worst-case regret optimizing* algorithms that set reserve prices in repeated second-price auctions with strategic bidders whose *valuation is private, but fixed* over all rounds. However, our setting constitutes a natural generalization of the well-studied 1-buyer setup of repeated posted-price auctions[1] (RPPA) (Amin et al., 2013; Mohri & Munoz, 2014) to the scenario of multiple buyers in a second-price auction. In the RPPA setting, there are optimal algorithms (Drutsa, 2017b; 2018) that have tight strategic regret bound of $\Theta(\log\log T)$. This bound follows from an ability of the seller to upper bound the buyer valuation even if he lies when rejecting a price (Drutsa, 2017b, Prop.2). This ability strongly exploits that the buyer knows in advance the outcomes of a current and all future rounds since he has complete information due to the absence of rivals. In our multi-bidder scenario, this does not hold: a bidder has incomplete information and is thus uncertain about the future. Hence, the theoretical guarantees could not be directly ported to our scenario when trying straightforwardly apply the optimal 1-buyer RPPA algorithms.

In our study, we propose a novel algorithm that can be applied against our strategic buyers with regret upper bound of $O(\log\log T)$ (Th. 1) and constitutes *the main contribution of our work*. We also introduce a novel transformation of a RPPA algorithm that maps it to a multi-buyer pricing and is based on a simple but crucial idea of cyclic elimination of all bidders except one in each round (Sec.3). Construction and

[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.

[1]In particular, when $M = 1$, our auction in a round reduces to a posted-price one: the bidder has no rivals and his decision is thus binary (to accept or to reject a currently offered price).

analysis of the proposed algorithm and transformation have required introduction of novel techniques, which are contributed by our work as well. They include (a) the method to locate the valuation of a strategic buyer in a played round under his uncertainty about the future (Prop. 1); (b) the decomposition of strategic regret into the regret of learning the individual valuations and the deviation regret of learning which bidder has the maximal valuation (Lemma 1); and (c) the approach to learn the highest-valuation bidder with deviation regret of $O(1)$ w.r.t. T (Lemma 3).

2. Preliminaries

2.1. Setup of Repeated Second-Price Auctions

We study the following mechanism of *repeated second-price auctions*. Namely, *the auctioneer* repeatedly proposes goods (e.g., advertisement opportunities) to M *bidders* (whose set is denoted by $\mathbb{M} := \{1, \ldots, M\}, M \in \mathbb{N}$) over T rounds: one good per round. From here on the following terminology is used as well: *the seller* for the auctioneer, a *buyer* for a bidder, and *the time horizon* for the number of rounds T. Each bidder $m \in \mathbb{M}$ holds a *fixed private valuation* $v^m \in [0, 1]$ for a good, i.e., the valuation v^m is equal for goods offered in all rounds and is unknown to the seller. The vector of valuations of all bidders is denoted by $\mathbf{v} := \{v^m\}_{m=1}^M$.

In each round $t \in \{1, \ldots, T\}$, for each bidder $m \in \mathbb{M}$, the seller sets a personal reserve price p_t^m, and the buyer m (knowing p_t^m) submits a sealed bid of b_t^m. Given the reserve prices $\mathbf{p}_t := \{p_t^m\}_{m=1}^M$ and the bids $\mathbf{b}_t := \{b_t^m\}_{m=1}^M$, the standard allocation and payment rules of a second price auction are applied (namely, the "eager" version (Paes Leme et al., 2016)): (a) for each bidder $m \in \mathbb{M}$, we check whether he bids over his reserve price or not, $a_t^m := \mathbb{I}_{\{b_t^m \geq p_t^m\}}$[2], obtaining the set $\mathbb{M}_t := \{m \in \mathbb{M} \mid a_t^m = 1\}$ of actual bidder-participants; (b) if $\mathbb{M}_t \neq \varnothing$, the good is allocated to the winning bidder $\overline{m}_t := \mathrm{argmax}_{m \in \mathbb{M}_t} b_t^m$ (if a tie, choose randomly) who pays $\overline{p}_t := \max\{p_t^{\overline{m}_t}, \max_{m \in \mathbb{M}_t \setminus \{\overline{m}_t\}} b_t^m\}$ to the seller. (c) if $\mathbb{M}_t = \varnothing$, the current good disappears and no payment is transferred. Further we use the following notations for allocation indicators, payments, and their vectors: $\overline{a}_t := \mathbb{I}_{\{\mathbb{M}_t \neq \varnothing\}}$, $\overline{a}_t^m := \mathbb{I}_{\{\mathbb{M}_t \neq \varnothing \& m = \overline{m}_t\}}$, $\overline{p}_t^m := \overline{a}_t^m \overline{p}_t$, $\mathbf{a}_t := \{a_t^m\}_{m=1}^M$, $\overline{\mathbf{a}}_t := \{\overline{a}_t^m\}_{m=1}^M$, and $\overline{\mathbf{p}}_t := \{\overline{p}_t^m\}_{m=1}^M$[3]. The summary on all notations is in App. C

Thus, the seller applies a *(pricing) algorithm* $\mathcal{A}$ that sets reserve prices $\mathbf{p}_{1:T} := \{\mathbf{p}_t\}_{t=1}^T$ in response to the buyers' bids $\mathbf{b}_{1:T} := \{\mathbf{b}_t\}_{t=1}^T$[4]. We consider the deterministic online learning case when the reserve price p_t^m for a bidder $m \in \mathbb{M}$ in a round $t \in \{1, \ldots, T\}$ can depend only on bids $\mathbf{b}_{1:t-1}$ of all bidders during the previous rounds and, possibly, the horizon T. Let $\mathbf{A}_M$ be the set of such algorithms. Hence, given a pricing algorithm $\mathcal{A} \in \mathbf{A}_M$, the buyers' bids $\mathbf{b}_{1:T}$ uniquely define the corresponding price sequence $\{\mathbf{p}_t\}_{t=1}^T$, which, in turn, determines the seller's total revenue $\sum_{t=1}^T \overline{a}_t \overline{p}_t$. This revenue is usually compared to the revenue that would have been earned by offering the highest valuation $\overline{v} := \max_{m \in \mathbb{M}} v^m$ if the valuations $\mathbf{v} = \{v^m\}_{m=1}^M$ were known in advance to the seller (Amin et al., 2013; Drutsa, 2017b). This leads to the notion of the *regret* of the algorithm $\mathcal{A}$:

$$\mathrm{Reg}(T, \mathcal{A}, \mathbf{v}, \mathbf{b}_{1:T}) := \sum_{t=1}^T (\overline{v} - \overline{a}_t \overline{p}_t).$$

Following a standard assumption in mechanism design that matches the practice in ad exchanges (Mohri & Munoz, 2014; Drutsa, 2018), the seller's pricing algorithm $\mathcal{A}$ *is announced to the buyers in advance*. A bidder can then act strategically against this algorithm. In contrast to the case of one bidder ($M = 1$), where the buyer can get an optimal behavior in advance, and the repeated mechanism reduces thus to a two-stage game (Amin et al., 2013; Mohri & Munoz, 2014; Drutsa, 2017b); in our setting, a bidder has incomplete information since he may not know the valuations and behaviors of the other bidders. Therefore, in order to model buyer strategic behavior under this uncertainty, we assume that, in each round t, each buyer optimizes his utility on subgame of future rounds given the available history of previous rounds and his beliefs about the other buyers.

Formally, in a round t, given the seller's pricing algorithm $\mathcal{A}$, a strategic buyer $m \in \mathbb{M}$ observes a history $h_t^m := (b_{1:t-1}^m, p_{1:t}^m, \overline{a}_{1:t-1}^m, \overline{p}_{1:t-1}^m)$ available to him and derives his *optimal bid* $\mathring{b}_t^m$ from a (possibly mixed) strategy $\sigma \in \mathfrak{S}_T$[5] that maximizes his future γ_m-discounted surplus:

$$\begin{aligned} &\mathrm{Sur}_{t:T}(\mathcal{A}, \gamma_m, v^m, h_t^m, \beta^m, \sigma) = \\ &= \mathbb{E}\Big[\sum_{s=t}^T \gamma_m^{s-1} \overline{a}_s^m (v^m - \overline{p}_s^m) \mid h_t^m, \sigma, \beta^m\Big], \end{aligned} \tag{1}$$

where $\gamma_m \in (0, 1]$ is *the discount rate*[6] of the bidder m. The expectation in Eq. (1) is taken over all possible continuations of the history h_t^m w.r.t. a strategy $\sigma \in \mathfrak{S}_T$ of the buyer m and his beliefs β^m about the strategies of the other bidders

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

[3] We use mnemonic notations: **boldface** for a vector over bidders and $\overline{\text{bar}}$ for terms associated with auction outcomes.

[4] $x_{t_1:t_2} = \{x_t\}_{t=t_1}^{t_2}$ denotes a part of a time series $\{x_t\}_{t=1}^T$.

[5] *A buyer strategy* is a map $\sigma : \mathbb{H}_{1:T} \to \mathbb{R}_+$ that maps any history $h \in \mathbb{H}_t$ in a round t to a bid $\sigma(h) \in \mathbb{R}_+$, where $\mathbb{H}_{1:T} := \sqcup_{t=1}^T \mathbb{H}_t$ and $\mathbb{H}_t := \mathbb{R}_+^{t-1} \times \mathbb{R}_+^t \times \mathbb{Z}_2^{t-1} \times \mathbb{R}_+^{t-1}$. Let $\mathfrak{S}_T$ denote the set of all possible strategies.

[6] Note that only buyer utilities are discounted over time, what is motivated by real-world markets as online advertising where sellers are far more willing to wait for revenue than buyers are willing to wait for goods (Mohri & Munoz, 2014; Drutsa, 2018).

$\mathbb{M}^{-m} := \mathbb{M} \setminus \{m\}$[7]. The buyer m assumes that the other bidders are strategic in the sense described above as well, what is taken into account in the beliefs β^m[8]. When T rounds has been played, let $\mathring{\mathbf{b}}_t := \{\mathring{b}_t^m\}_{m=1}^M$ be the optimal bids that depend on $(T, \mathcal{A}, \mathbf{v}, \boldsymbol{\gamma}, \boldsymbol{\beta})$, where $\boldsymbol{\gamma} = \{\gamma_m\}_{m=1}^M$ and $\boldsymbol{\beta} = \{\beta_m\}_{m=1}^M$. We define *the strategic regret* of the algorithm $\mathcal{A}$ that faced M strategic buyers with valuations $\mathbf{v} \in [0,1]^M$ and beliefs $\boldsymbol{\beta}$ over T rounds as

$$\mathrm{SReg}(T, \mathcal{A}, \mathbf{v}, \boldsymbol{\gamma}, \boldsymbol{\beta}) := \mathrm{Reg}\big(T, \mathcal{A}, \mathbf{v}, \mathring{\mathbf{b}}_{1:T}(T, \mathcal{A}, \mathbf{v}, \boldsymbol{\gamma}, \boldsymbol{\beta})\big).$$

In our setting, following (Amin et al., 2013; Mohri & Munoz, 2014; Drutsa, 2017b; 2018), we seek for algorithms that attain $o(T)$ strategic regret for the *worst-case* valuations $\mathbf{v} \in [0,1]^M$. Formally, an algorithm $\mathcal{A}$ is said to be a *no-regret* one when $\sup_{\mathbf{v} \in [0,1]^M, \boldsymbol{\beta}} \mathrm{SReg}(T, \mathcal{A}, \mathbf{v}, \boldsymbol{\gamma}, \boldsymbol{\beta}) = o(T)$ in our multi-buyer case. The optimization goal is to find algorithms with the lowest possible strategic regret upper bound $O(f(T))$, i.e., $f(T)$ has the slowest growth as $T \to \infty$ or, alternatively, the averaged regret has the best rate of convergence to zero.

2.2. Background on Pricing Algorithms

To the best of our knowledge, there is no work studied *worst-case regret optimizing* algorithms that set reserve prices in repeated second-price auctions with strategic bidders whose *valuation is private, but fixed* over all rounds. However, in the case of one bidder, $M = 1$, the bidder has no rivals, and, thus, the second-price auction in a round t reduces to a posted-price auction, where the buyer decision reduces to a binary action: to accept or to reject a currently offered price p_t^1. Let $\mathbf{A}^{\mathrm{RPPA}} \subset \mathbf{A}_1$ be the subclass of the 1-bidder algorithms s.t. each reserve price p_t^1 depends only on the past binary decisions $a_{1:t-1}^1$ of the buyer to get or do not get a good for a posted reserve price. For this subclass, all our strategic setting of repeated second-price auctions reduces to the setup of repeated posted-price auctions (RPPA) earlier introduced in (Amin et al., 2013).

Pricing algorithms in the strategic setup of RPPA with fixed private valuation and worst-case regret optimization were well studied last years (Amin et al., 2013; Mohri & Munoz, 2014; Drutsa, 2017b; 2018). It is known that, if the discount rate $\gamma = 1$, any algorithm has a linear strategic regret, i.e., the regret has lower bound $\Omega(T)$ (Amin et al., 2013), while, for the other cases $\gamma \in (0,1)$, the lower bound of $\Omega(\log\log T)$ holds (Kleinberg & Leighton, 2003; Mohri & Munoz, 2014). The first algorithm with optimal strategic regret bound of $\Theta(\log\log T)$ was found in (Drutsa, 2017b). It is Penalized Reject-Revising Fast Exploiting Search (PRRFES), which is horizon-independent and is based on Fast Search (Kleinberg & Leighton, 2003) modified to act against a strategic buyer. The modifications include penalizations (see Def. 1). A strategic buyer either accepts the price at the first node or rejects this price in subsequent penalization ones (Mohri & Munoz, 2014; Drutsa, 2017b). PRRFES is also a right-consistent algorithm: a RPPA algorithm $\mathcal{A}_1$ is *right-consistent* ($\mathcal{A}_1 \in \mathbf{C_R}$) if it never offers a price lower than the last accepted one (Drutsa, 2017b). The pricing algorithm PRRFES was further modified by the transformation `pre` to obtain the one that never decreases offered prices and has a tight strategic regret bound of $\Theta(\log\log T)$ as well (Drutsa, 2018).

The workflow of a RPPA algorithm $\mathcal{A}_1$ is usually described by a labeled binary tree $\mathfrak{T}(\mathcal{A}_1)$ (Mohri & Munoz, 2014; Drutsa, 2017b; 2018): initialize the tracking node $\mathfrak{n}$ to the root $\mathfrak{e}(\mathfrak{T}(\mathcal{A}_1))$; in each round, the label $\mathrm{p}(\mathfrak{n})$ is offered as a price; if it is accepted (rejected), move the tracking node to the right child $\mathfrak{n} := \mathfrak{r}(\mathfrak{n})$ (the left child $\mathfrak{n} := \mathfrak{l}(\mathfrak{n})$, resp.); and go to the next round. The left (right) subtrees rooted at the node $\mathfrak{l}(\mathfrak{n})$ ($\mathfrak{r}(\mathfrak{n})$, resp.) are denoted by $\mathfrak{L}(\mathfrak{n})$ ($\mathfrak{R}(\mathfrak{n})$, resp.). When trees $\mathfrak{T}_1$ and $\mathfrak{T}_2$ have the same node labeling, we write $\mathfrak{T}_1 \cong \mathfrak{T}_2$.

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

2.3. Overview of our Results

We cannot directly apply the optimal RPPA algorithms (Drutsa, 2017b; 2018), because our bidders have incomplete information in the game, while the proofs of optimality of these algorithms strongly rely on complete information. This completely different information structure of the multi-buyer game results in very complicated bidder behavior even in the absence of reserve prices (Bikhchandani, 1988). Hence, it is challenging to find, in the multi-buyer case, a pricing algorithm that has regret upper bound of the same asymptotic behavior as the best one in the 1-buyer RPPA setting. Our research goal comprises closing of this research question on the existence of such algorithms.

First, we propose a novel technique to transform a RPPA algorithm to our setup that is based on cyclic elimination of all bidders except one by means of high enough prices (Sec. 3). Separate playing with each buyer removes his uncertainty about the outcome of a current round; and, despite remaining uncertainty about future rounds, this is enough to construct a tool to locate his valuation (Prop. 1). Second, we transform PRRFES in this way and show that its regret is affected by two learning processes: the one learns bidder valuations and the other learns which bidders have the max-

[7] So, σ and β^m determine the future outcomes $\overline{a}_s^m$ and $\overline{p}_s^m$, that are thus random variables.

[8] In our setup, we do not require that the strategies actually used by the buyers $\mathbb{M}^{-m}$ match with the buyer m's beliefs β^m (an equilibrium requirement), because our results hold without this requirement.

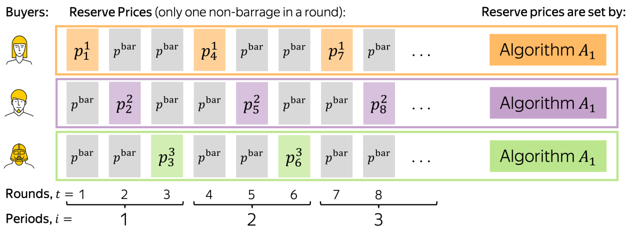

Figure 1. An illustration of cyclic elimination of bidders by means of a barrage price in a dividing algorithm. In this example, there are $M = 3$ bidders, three periods are depicted, and the reserve prices are set independently for the bidders by a RPPA algorithm A_1.

imal valuation (Sec. 4). The former learning is controlled by the design of the source PRRFES, while the latter one is achieved by a special stopping rule that excludes bidders from suspected ones. A proper combination of parameters for the source pricing and the stopping rule provides an algorithm with strategic regret in $O(\log\log T)$, see Th. 1.

2.4. Related Work

Several studies maximized revenue of auctions in an offline/batch learning fashion: either via estimating or fitting of distributions of buyer valuations/bids to set reserve prices (He et al., 2013; Sun et al., 2014; Paes Leme et al., 2016), or via direct learning of reserve prices (Mohri & Medina, 2014; 2015; Rudolph et al., 2016; Medina & Vassilvitskii, 2017). In contrast to them, we set prices in repeated auctions by an online deterministic learning approach.

Revenue optimization for repeated auctions was mainly concentrated on algorithmic reserve prices, that are updated in online way over time, and was also known as dynamic pricing (Fudenberg & Villas-Boas, 2006; den Boer, 2015). Dynamic pricing was considered: under game-theoretic view (Leme et al., 2012; Chen & Farias, 2015; Balseiro et al., 2016; Ashlagi et al., 2016; Mirrokni et al., 2018; Abernethy et al., 2019); from the bidder side (Iyer et al., 2011; Weed et al., 2016; Heidari et al., 2016; Baltaoglu et al., 2017); in experimental studies (List & Shogren, 1999; Carare, 2012; Yuan et al., 2014); as bandit problems (Amin et al., 2011; Lin et al., 2015; Cesa-Bianchi et al., 2018); and from other aspects (Roughgarden & Wang, 2016; Feldman et al., 2016; Chawla et al., 2016; Hummel, 2018; Deng et al., 2019b). Repeated auctions with a contextual information about the good in a round were considered in (Amin et al., 2014; Cohen et al., 2016; Mao et al., 2018; Leme & Schneider, 2018; Golrezaei et al., 2019; Deng et al., 2019a; Drutsa, 2020; Zhiyanov & Drutsa, 2020). The studies (Schmidt, 1993; Hart & Tirole, 1988; Devanur et al., 2015; Immorlica et al., 2017; Vanunts & Drutsa, 2019) elaborated on setups of repeated posted-price auctions with a strategic buyer holding a fixed valuation, but maximized expected revenue for a given prior distribution of valuations, while we optimize regret w.r.t. worst-case valuations without knowing their distribution.

There are studies on reserve price optimization in repeated second-price auctions, but they considered scenarios different to ours. Non-strategic bidders are considered in (Cesa-Bianchi et al., 2013). Kanoria et al. (Kanoria & Nazerzadeh, 2014) studied strategic buyers (similarly to our work), but maximized expected revenue w.r.t. a prior distribution of valuations. Our setup can be considered as a special case of repeated Vickrey auctions in (Huang et al., 2018), but their regret upper bound is $O(T^{\alpha})$ in T and holds only when selling several goods in a round. However, the most relevant works to ours are (Amin et al., 2013; Mohri & Munoz, 2014; Drutsa, 2017b; 2018), where our strategic setup with fixed private valuation is considered, but for the case of one bidder, $M = 1$. The most important results of these works are discussed in Sec. 2.2.

3. Dividing Algorithm and div-transformation

3.1. Barrage Pricing

In our setting, a pricing algorithm is able to set personal (individual) reserve prices to each bidder and, hence, is able to "eliminate" particular bidders from particular rounds. Namely, in a round t, an algorithm can set a reserve price p^{bar} s.t. a strategic bidder m, independently of his valuation, will never accept p^{bar}, i.e., will never bid higher than this price; such a price is referred to as a *barrage reserve price*. From here on we use $p^{\mathrm{bar}} = 1/(1-\gamma_0)$, $\gamma_0 \in (0,1)$: accepting it once will result in a negative surplus for a buyer with discount $\gamma_m \le \gamma_0$. We use the phrase "the bidder m is *eliminated*[9] from participation in the round t" to describe this case.

3.2. Dividing Algorithms

In this subsection, we introduce a subclass of the algorithms $\mathbf{A}_M$ that is denoted by $\mathbf{A}_M^{\mathrm{div}} \subset \mathbf{A}_M$ and is referred to as *the class of dividing algorithms* (stands for lat. "Divide et impera"). A dividing algorithm $\mathcal{A} \in \mathbf{A}_M^{\mathrm{div}}$ works in *periods* and tracks a feasible set of *suspected bidders* $\mathbb{S}$ aimed to find the bidder (or bidders) with the maximal valuation $\overline{v}$. Namely, it starts with all bidders $\mathbb{S}_1 := \mathbb{M}$ at the first period which lasts M rounds. In each period $i \in \mathbb{N}$, the algorithm iterates over the currently suspected bidders $\mathbb{S}_i$: in a current round, it picks up $m \in \mathbb{S}_i$, sets a non-barrage reserve price to the bidder m, sets a barrage reserve price to all other bidders $\mathbb{M}^{-m}$, and goes to the next round within the period

[9] Note that, (a) formally, all bidders participate in all rounds (see Sec. 2) and (b), if a bidder is not eliminated, it does not mean that he is in $\mathbb{M}_t$ (he may bid below his reserve price which can be a non-barrage one). So, the word "elimination" is purposely associate with barrage pricing in order to refer to this case.

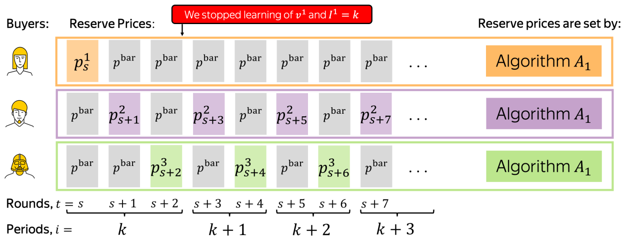

Figure 2. An illustration of application of a stopping rule by means of a barrage price in a dividing algorithm. In this example, there are $M = 3$ bidders, four periods are depicted, and the buyer 1 is no longer considered after the k-th period (his subhorizon is $I^1 = k$).

by picking up the next buyer from $\mathbb{S}_i$. Thus, the algorithm meaningfully interacts with only one bidder in each round through elimination of all other bidders by means of barrage pricing. After the i-th period, the algorithm $\mathcal{A}$ identifies somehow which bidders from $\mathbb{S}_i$ should be left as suspected ones in the next period (i.e., be included in the set $\mathbb{S}_{i+1}$). The processes of cyclic elimination and stopping rule are illustrated in Figures 1 and 2, respectively.

When the game has been played with the dividing algorithm $\mathcal{A}$, one can split all the rounds into I periods: $\{1, \ldots, T\} = \cup_{i=1}^{I} \mathcal{T}_i$. Each period $i < I$ consists of $|\mathcal{T}_i| = |\mathbb{S}_i|$ rounds (the last one of $|\mathcal{T}_I| \leq |\mathbb{S}_I|$). Let $t_i^m \in \mathcal{T}_i$ denote the round of a period i in which a bidder m is not eliminated by the seller algorithm (i.e., receives a non-barrage reserve price). Thus, $\mathcal{I}^m := \{t_1^m, ..., t_{I^m}^m\}$ are all such rounds of the bidder m and $I^m = |\mathcal{I}^m|$ is referred to as the *subhorizon* of the bidder m (the number of periods where he participates). Note that (a) I^m and $\mathcal{I}^m$ depend on the bids $\mathbf{b}_{1:T}$ of all buyers $\mathbb{M}$; (b) the following identities hold: $\{1, \ldots, T\} = \cup_{m=1}^{M} \mathcal{I}^m$ and $\mathcal{T}_i = \{t_i^m \mid m \in \mathbb{M} \text{ s.t. } I^m \geq i\}$.

So, in a round t_i^m, the algorithm $\mathcal{A}$ eliminates the bidders $\mathbb{M}^{-m}$ (i.e., sets the reserves $p_{t_i^m}^{m'} = p^{\text{bar}}\ \forall m' \in \mathbb{M}^{-m}$), while the reserve price $p_{t_i^m}^m$ set for the buyer m is determined only by his bids during the previous rounds $\{t_1^m, ..., t_{i-1}^m\}$ where he has not been eliminated: i.e., $p_{t_i^m}^m = p^m(b_{t_1^m}^m, ..., b_{t_{i-1}^m}^m)$. Hence, the algorithm $\mathcal{A}$'s interaction with the bidder m in the rounds $\mathcal{I}^m$ can be encoded by a 1-buyer algorithm from $\mathbf{A}_1$, which sets prices in the rounds $\{t_i^m\}_{i=1}^{I^m}$ instead of $\{i\}_{i=1}^{I^m}$. We denote this algorithm by $\mathcal{A}^m$ and refer to it as *the subalgorithm* of $\mathcal{A}$ against the buyer m. Let

$$\mathrm{Reg}^m(\mathcal{I}^m, \mathcal{A}^m, v^m, b_{1:T}^m) := \sum_{i=1}^{I^m} (v^m - a_{t_i^m}^m p_{t_i^m}^m)$$

be the regret of the subalgorithm $\mathcal{A}^m$ for given bids $b_{1:T}^m$ of the buyer $m \in \mathbb{M}$ in the rounds $\mathcal{I}^m$. The following lemma holds (the trivial proof is in Appendix A.1.1 in Supp.Mat.).

Lemma 1. *Let* $\mathcal{A} \in \mathbf{A}_M^{\text{div}}$ *be a dividing algorithm,* $\mathcal{A}^m \in \mathbf{A}_1, m \in \mathbb{M}$*, be its subalgorithms (as described above), and* $\mathring{\mathbf{b}}_{1:T} = \mathring{\mathbf{b}}_{1:T}(T, \mathcal{A}, \mathbf{v}, \boldsymbol{\gamma}, \boldsymbol{\beta})$ *be optimal bids of the strategic buyers* $\mathbb{M}$*. Then, for any* $\mathbf{v} \in [0,1]^M$*,* $\boldsymbol{\gamma} \in (0,1]^M$*, and* $\boldsymbol{\beta}$*, the strategic regret of* $\mathcal{A}$ *can be decomposed into two parts*

$$\mathrm{SReg}(T, \mathcal{A}, \mathbf{v}, \boldsymbol{\gamma}, \boldsymbol{\beta}) = \mathrm{SReg}^{\text{ind}}(T, \mathcal{A}, \mathbf{v}, \boldsymbol{\gamma}, \boldsymbol{\beta}) + \\ + \mathrm{SReg}^{\text{dev}}(T, \mathcal{A}, \mathbf{v}, \boldsymbol{\gamma}, \boldsymbol{\beta}),$$

where

$$\mathrm{SReg}^{\text{ind}}(T, \mathcal{A}, \mathbf{v}, \boldsymbol{\gamma}, \boldsymbol{\beta}) := \sum_{m \in \mathbb{M}} \mathrm{Reg}^m(\mathcal{I}^m, \mathcal{A}^m, v^m, \mathring{b}_{1:T}^m)$$

is the individual part of the regret and

$$\mathrm{SReg}^{\text{dev}}(T, \mathcal{A}, \mathbf{v}, \boldsymbol{\gamma}, \boldsymbol{\beta}) := \sum_{m \in \mathbb{M}} I^m(\overline{v} - v^m)$$

is the deviation part of the regret.

Informally, this lemma states that the regret consists of the individual regrets against each buyer m in his rounds $\mathcal{I}^m$ and the deviation of the buyer valuations $\mathbf{v}$ from the highest valuation $\overline{v}$. So, we see a clear intuition: a good algorithm should (1) *learn the valuations* $\mathbf{v}$ *of the buyers* (minimizing individual regrets) and (2) *learn which buyers have the highest valuation* $\overline{v}$ (minimizing the deviation regret).

3.3. div-transformation

Let $\mathcal{A}_1 \in \mathbf{A}^{\text{RPPA}}$ be a 1-buyer RPPA algorithm. A M-buyer algorithm $\mathrm{div}_M(\mathcal{A}_1, \mathrm{sr})$ is said to be *a div-transformation of the algorithm* $\mathcal{A}_1$ with *a stopping rule* $\mathrm{sr} : \mathbb{M} \times \mathfrak{T}(\mathcal{A}_1)^M \to \texttt{bool}$ when it is a dividing algorithm from $\mathbf{A}_M^{\text{div}}$ s.t. its subalgorithms $\mathcal{A}^m$ are $\mathcal{A}_1$ and the stopping rule sr determines which bidders are not suspected after the end of each period i (i.e., which bidders do not present in the set $\mathbb{S}_{i+1}$). Namely, first, the algorithm $\mathrm{div}_M(\mathcal{A}_1, \mathrm{sr})$ tracks the state of each buyer $m \in \mathbb{M}$ in the tree $\mathfrak{T}(\mathcal{A}_1)$ of the RPPA algorithm $\mathcal{A}_1$ (see Sec. 2) by means of a personal (individual) feasible node: for each period i and for each round $t_i^m \in \mathcal{T}_i$, the current state (i.e., the history of previous actions) of the buyer m is encoded by the tracking node $\mathfrak{n}_i^m \in \mathfrak{T}(\mathcal{A})$. In particular, in the round t_i^m, the buyer m receives the reserve price equal to the price $\mathrm{p}(\mathfrak{n}_i^m)$ of this node $\mathfrak{n}_i^m$ (the other bidders $\mathbb{M}^{-m}$ get a barrage reserve price p^{bar}). If a buyer m is not more suspected in a period $i > I^m$ (i.e., $m \notin \mathbb{S}_i$), we formally set $\mathfrak{n}_i^m := \mathfrak{n}_{I^m+1}^m$. Second, after a period i, the stopping decision for a buyer m' is based on the past buyer binary actions that are coded by means of the nodes $\{\mathfrak{n}_{i+1}^m\}_{m=1}^M$ in the binary tree $\mathfrak{T}(\mathcal{A}_1)$: if the stopping rule $\mathrm{sr}(m', \{\mathfrak{n}_{i+1}^m\}_{m=1}^M)$ is $\texttt{true}$, then the buyer $m' \notin \mathbb{S}_{i+1}$. The pseudo-code of the div-transformation of a RPPA algorithm is in Algorithm 1.

For a RPPA right-consistent algorithm $\mathcal{A}_1 \in \mathbf{C_R}$ with penalization rounds, let $\langle \mathcal{A}_1 \rangle$ denote the transformation of $\mathcal{A}_1$ s.t. it is equal to $\mathcal{A}_1$, but each penalization sequence

Algorithm 1 Pseudo-code of a div-transformation $\mathrm{div}_M(\mathcal{A}_1, \mathrm{sr})$ of a RPPA algorithm $\mathcal{A}_1 \in \mathbf{A}^{\mathrm{RPPA}}$.

Input: $M \in \mathbb{N}$, $\mathcal{A}_1 \in \mathbf{A}^{\mathrm{RPPA}}$, $\mathrm{sr} : \mathbb{M} \times \mathfrak{T}(\mathcal{A}_1)^M \to \texttt{bool}$
Initialize: $t := 1, \mathbb{S} := \mathbb{M}, \mathfrak{n}[\,] := \{\mathfrak{e}(\mathfrak{T}(\mathcal{A}_1))\}_{m=1}^{M}$
while $t \le T$ **do**
  for all $m \in \mathbb{S}$ **do**
    Set the price $\mathrm{p}(\mathfrak{n}[m])$ as reserve to the buyer m
    Set the price p^{bar} as reserve to the buyers from $\mathbb{M}^{-m}$
    $\mathbf{b}[\,] \leftarrow$ get bids from the buyers $\mathbb{M}$
    if $\mathbf{b}[m] \ge \mathrm{p}(\mathfrak{n}[m])$ **then**
      Allocate t-th good to the buyer m for the price $\mathrm{p}(\mathfrak{n}[m])$
      $\mathfrak{n}[m] := \mathfrak{r}(\mathfrak{n}[m])$
    else
      $\mathfrak{n}[m] := \mathfrak{l}(\mathfrak{n}[m])$
    end if
    $t := t + 1$
    if $t > T$ **then**
      break
    end if
  end for
  $\mathbb{S}^{\mathrm{old}} := \mathbb{S}$
  for all $m \in \mathbb{S}^{\mathrm{old}}$ **do**
    if $\mathrm{sr}(m, \mathfrak{n}[\,])$ **then**
      $\mathbb{S} := \mathbb{S} \setminus \{m\}$
    end if
  end for
end while

of nodes $\{\mathfrak{n}_j\}_{j=1}^r \subset \mathfrak{T}(\mathcal{A}_1), r \ge 2$, (see Def. 1) is *reinforced* in the following way: all the prices in the nodes $\{\mathfrak{n}_j\} \cup \mathfrak{R}(\mathfrak{n}_j), j = 2, ..., r$, are replaced by 1 (the maximal valuation domain value); the sequence and the rounds are then referred to as *reinforced penalization* ones. After this, a strategic buyer will certainly either accept the price at the node $\mathfrak{n}_1$, or reject the prices in all the nodes $\{\mathfrak{n}_j\}_{j=1}^r$ even in the case of his uncertainty about the future. Let $\delta^l_{\mathfrak{n}} := \mathrm{p}(\mathfrak{n}) - \inf_{\mathfrak{m} \in \mathfrak{L}(\mathfrak{n})} \mathrm{p}(\mathfrak{m})$ be the left increment (Mohri & Munoz, 2014; Drutsa, 2017b) of a node $\mathfrak{n} \in \mathfrak{T}(\mathcal{A}_1)$.

In order to obtain upper bounds on strategic regret, it is important to have a tool that allows to locate the valuation of a strategic bidder. Such a tool can be obtained for div-transformed right-consistent RPPA algorithms with reinforced penalization rounds based on the following proposition, which is an analogue of (Drutsa, 2017b, Prop.2) in our case with buyer uncertainty about the future.

Proposition 1. *Let* $\gamma_m \in (0, 1)$, $\mathcal{A}_1 \in \mathbf{A}^{\mathrm{RPPA}} \cap \mathbf{C_R}$ *be a RPPA right-consistent pricing algorithm,* $\mathfrak{n} \in \mathfrak{T}(\mathcal{A}_1)$ *be a starting node in a* r*-length penalization sequence (see Def. 1),* $r > \log_{\gamma_m}(1 - \gamma_m)$, $\mathrm{sr} : \mathbb{M} \times \mathfrak{T}(\mathcal{A}_1)^M \to \texttt{bool}$ *be a stopping rule, and the* div*-transformation* $\mathrm{div}_M(\langle \mathcal{A}_1 \rangle, \mathrm{sr})$ *be used by the seller for setting reserve prices. If, in a round, the node* $\mathfrak{n}$ *is reached and the price* $\mathrm{p}(\mathfrak{n})$ *is rejected by a strategic buyer* $m \in \mathbb{M}$ *(i.e., he bids lower than* $\mathrm{p}(\mathfrak{n})$*), then the following inequality on* v^m *holds:*

$$v^m - \mathrm{p}(\mathfrak{n}) < \zeta_{r,\gamma_m} \delta^l_{\mathfrak{n}}, \;\; \textit{where} \;\; \zeta_{r,\gamma} := \frac{\gamma^r}{1 - \gamma - \gamma^r}. \quad (2)$$

Proof sketch. The full proof is in App.A.1.2. Let t be the round in which the bidder m reaches the node $\mathfrak{n}$ and rejects his reserve price $p^m_t = \mathrm{p}(\mathfrak{n})$. In particular, it is the round where he is the non-eliminated buyer and $t = t^m_i \in \mathcal{T}_i$ for some period i. Since the buyers are divided and $\mathcal{A}_1 \in \mathbf{A}^{\mathrm{RPPA}}$, w.l.o.g., any strategy can be treated as a map to binary decisions $\{0, 1\}$. Let $\mathring{\sigma}$ be the optimal strategy used by the buyer m; $h^m_{t;a}$ be the continuation of the current history h^m_t by a binary decision $a^m_t = a$, while $\hat{\sigma}_a$ denote an optimal strategy among all possible strategies in which the binary buyer decision a^m_t is $a \in \{0, 1\}$; and

$$S^m_t(\sigma) := \mathrm{Sur}_{t:T}(\mathcal{A}, \gamma_m, v^m, h^m_t, \beta^m, \sigma)$$

be the future expected surplus when following a strategy $\sigma \in \mathfrak{S}_T$. Rejection of the price p^m_t when following the optimal strategy $\mathring{\sigma}$ easily implies: $S^m_t(\hat{\sigma}_1) \le S^m_t(\hat{\sigma}_0)$. Let us bound each side of this inequality. First,

$$\begin{aligned} S^m_t(\hat{\sigma}_1) = \gamma_m^{t-1}(v^m - \mathrm{p}(\mathfrak{n})) + \\ + \mathrm{Sur}_{t+1:T}(\mathcal{A}, \gamma_m, v^m, h^m_{t;1}, \beta^m, \hat{\sigma}_1) \ge \\ \ge \gamma_m^{t-1}(v^m - \mathrm{p}(\mathfrak{n})), \end{aligned} \quad (3)$$

where we used the facts (i) that if the bidder accepts the price $\mathrm{p}(\mathfrak{n})$, then he necessarily gets the good since all other bidders $\mathbb{M}^{-m}$ are eliminated by a barrage price in this round t; and (ii) that the expected surplus in rounds $s \ge t+1$ is at least non-negative, because the subalgorithm $\mathcal{A}_1 \in \mathbf{C_R}$ is right-consistent. Second,

$$\begin{aligned} S^m_t(\hat{\sigma}_0) = \mathrm{Sur}_{t^m_{i+r}:T}(\mathcal{A}, \gamma_m, v^m, h^m_{t;0}, \beta^m, \hat{\sigma}_0) < \\ < \frac{\gamma_m^{t+r-1}}{1 - \gamma_m}(v^m - \mathrm{p}(\mathfrak{n}) + \delta^l_{\mathfrak{n}}), \end{aligned} \quad (4)$$

where we (i) used the fact that if the bidder rejects the price p^m_t, then the future rounds $\{t^m_{i+j}\}_{j=1}^{r-1}$ will be reinforced penalization ones (the strategic bidder will reject in all of them); and (ii) upper bounded the surplus in remaining rounds by assuming that only this bidder will get remaining goods for the lowest reserve price from the left subtree $\mathfrak{L}(\mathfrak{n})$. We unite these bounds on $S^m_t(\hat{\sigma}_a)$, divide by γ_m^{t-1}, and get

$$(v^m - \mathrm{p}(\mathfrak{n}))\left(1 - \frac{\gamma_m^r}{1 - \gamma_m}\right) < \frac{\gamma_m^r}{1 - \gamma_m}\delta^l_{\mathfrak{n}}, \quad (5)$$

what implies Eq. (2), since $r > \log_{\gamma_m}(1 - \gamma_m)$. □

We emphasize that *the dividing structure of the algorithm is crucially exploited* in the proof of Prop. 1. Namely, the fact that all other bidders $\mathbb{M}^{-m}$ are eliminated by a barrage price in the round t is used (a) to guarantee obtaining of the good at price $\mathrm{p}(\mathfrak{n})$ by the buyer m and (b) to lower bound thus the future surplus $S^m_t(\hat{\sigma}_1)$ in the case of acceptance in Eq. (3). If we dealt with a non-dividing algorithm, then another bidder might win the good or make the payment of the bidder m higher than his reserve price $\mathrm{p}(\mathfrak{n})$; in both cases, $S_t^m(\hat{\sigma}_1)$ could only be lower bounded by 0 in a general situation, what would result in an useless inequality instead of Eq. (2).

For a right-consistent algorithm $\mathcal{A}_1 \in \mathbf{C_R}$, the increment $\delta_{\mathfrak{n}}^l$ is bounded by the difference between the current node's price $\mathrm{p}(\mathfrak{n})$ and the last price q that has been accepted by the buyer m before reaching this node. Hence, the Prop. 1 provides us with a tool to locate the valuation v^m despite the strategic buyer does not myopically report its position (similar to (Drutsa, 2017b, Prop.2)). Namely, if the buyer m bids no lower than $\mathrm{p}(\mathfrak{n})$, then $v^m \geq \mathrm{p}(\mathfrak{n})$; if he bids lower than $\mathrm{p}(\mathfrak{n})$, then $q \leq v < \mathrm{p}(\mathfrak{n}) + \zeta_{r,\gamma_m}(\mathrm{p}(\mathfrak{n}) - q)$ and the closer an offered price $\mathrm{p}(\mathfrak{n})$ is to the last accepted price q the smaller the location interval of possible valuations v^m (since its length is $(1+\zeta_{r,\gamma_m})(\mathrm{p}(\mathfrak{n}) - q)$).

4. divPRRFES Algorithm

In this section, we will show that we can use an optimal algorithm from the setting of repeated posted-price auctions to obtain the algorithm for our multi-bidder setting with upper bound on strategic regret with the same asymptotic. Namely, let us div-transform PRRFES (Drutsa, 2017b), further denoted as $\mathcal{A}_1$.

Since a div-transformation of PRRFES (with penalization reinforcement) individually tracks position of each buyer in the binary tree $\mathfrak{T}(\langle\mathcal{A}_1\rangle)$, we adapt the key notations of PRRFES (Drutsa, 2017b) to our case of multiple bidders and periods. Against a buyer $m \in \mathbb{M}$, PRRFES $\langle\mathcal{A}_1\rangle$ works in phases initialized by the phase index $l := 0$, the last accepted price before the current phase $q_0^m := 0$, and the iteration parameter $\epsilon_0 := 1/2$. At each phase $l \in \mathbb{Z}_+$, it sequentially offers prices $p_{l,k}^m := q_l^m + k\epsilon_l, k \in \mathbb{N}$ (*exploration rounds*), with $\epsilon_l = 2^{-2^l}$; if a price $p_{l,k}^m$ is rejected, setting $K_l^m := k - 1 \geq 0$,

1. it offers the price 1 for $r-1$ *reinforced penalization rounds* (if one of them is accepted, 1 will be offered in all remaining rounds),
2. it offers the price $p_{l,K_l^m}^m$ for $g(l)$ *exploitation rounds*,
3. PRRFES goes to the next phase by setting $q_{l+1}^m := p_{l,K_l^m}^m$ and $l := l + 1$. Individual tracking of bidders by the div-transformed PRRFES implies that different buyers can be in different phases in the same period i.

Hence, let l_i^m denote the current phase of a buyer $m \in \mathbb{M}$ in the round t_i^m of a period $i \leq I^m$, and let $l_i^m := l_{I^m+1}^m$ in all subsequent periods $i > I^m$ (when the buyer m is no more suspected). In particular, $q_{l_i^m}^m$ is the last accepted price by the buyer m before the phase l_i^m in the period i.

We rely on the decomposition from Lemma 1 in order to bound the strategic regret of a div-transformed PRRFES.

4.1. Upper Bound for Individual Regrets

Before specifying a particular stopping rule, let us obtain an upper bound on individual strategic regret $\mathrm{Reg}^m(\mathcal{I}^m, \langle\mathcal{A}_1\rangle, v^m, \mathring{b}_{1:T}^m), m \in \mathbb{M}$. This regret is not equal to $\mathrm{SReg}(I^m, \langle\mathcal{A}_1\rangle, (v^m), (\gamma_m))$ since, in the latter case, the 1-bidder game does not depend on behavior of the other bidders $\mathbb{M}^{-m}$ (while, in the former case, does). In other words, the rounds $\mathcal{I}^m = \{t_i^m\}_{i=1}^{I^m}$ do not constitute the I^m-round 1-buyer game of the RPPA setting considered in (Amin et al., 2013; Drutsa, 2017b), because the subhorizon I^m and exact rounds $\mathcal{I}^m$ (they determine the used discount factors: $\gamma_m^{t-1}, t \in \mathcal{I}^m$) are unknown in advance and depend on actions of the other bidders. Hence, this does not allow to straightforwardly utilize the result on the strategic regret for PRRFES proved in (Drutsa, 2017b, Th.5) for the setting of RPPA. So, we have to prove the bound $O(\log\log T)$ for our case with buyer uncertainty about the future. Let us introduce the notation:

$$r_\gamma := \left\lceil \log_\gamma \left((1-\gamma)/2\right)\right\rceil \; \forall \gamma \in (0,1). \tag{6}$$

Lemma 2. *Let $\gamma_0 \in (0,1)$, $\mathcal{A}_1$ be the PRRFES algorithm with $r \geq r_{\gamma_0}$ and the exploitation rate $g(l) = 2^{2^l}, l \in \mathbb{Z}_+$, and $\mathrm{sr} : \mathbb{M}\times\mathfrak{T}(\mathcal{A}_1)^M \to$ `bool` be a stopping rule. Then, for any valuation $v^m \in [0,1]$, if $I^m \geq 2$, the individual regret of the div-transformed PRRFES $\mathrm{div}_M(\langle\mathcal{A}_1\rangle, \mathrm{sr})$ against the buyer $m \in \mathbb{M}$ is upper bounded:*

$$\begin{aligned}&\mathrm{Reg}^m(\mathcal{I}^m, \langle\mathcal{A}_1\rangle, v^m, \mathring{b}_{1:T}^m) \leq \\ &\leq (rv^m + 4)(\log_2\log_2 I^m + 2) \quad \forall \gamma_m \in (0, \gamma_0]\end{aligned} \tag{7}$$

where $\mathring{\mathbf{b}}_{1:T} = \mathring{\mathbf{b}}_{1:T}(T, \mathrm{div}_M(\langle\mathcal{A}_1\rangle, \mathrm{sr}), \mathbf{v}, \boldsymbol{\gamma}, \boldsymbol{\beta})$ are optimal bids of the strategic buyers $\mathbb{M}$.

Proof sketch. Decompose the individual regret over the rounds $\mathcal{I}^m$ into the sum of the phases' regrets:

$$\mathrm{Reg}^m(\mathcal{I}^m, \langle\mathcal{A}_1\rangle, v^m, \mathring{b}_{1:T}^m) = \sum_{l=0}^{L^m} R_l^m,$$

where $L^m := l_{I^m}^m$ is the number of phases conducted by the algorithm against the buyer m. For $l \in \mathbb{Z}_{L^m-1}$:

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

where the terms correspond to the accepted exploration rounds, the reject-penalization ones, and the exploitation ones. PRRFES and each rejected price $p_{l,K_l^m+1}^m$ satisfy the conditions of Prop. 1, what implies

$$v^m - p_{l,K_l^m+1}^m < (p_{l,K_l^m+1}^m - p_{l,K_l^m}^m) = \epsilon_l$$

since $\zeta_{r,\gamma_m} \leq 1$ for $r \geq r_{\gamma_0}$ and $\gamma_m \leq \gamma_0$. Hence, $v^m \in [q^m_{l+1}, q^m_{l+1} + 2\epsilon_l)$ (since $q^m_{l+1} = p^m_{l,K^m_l}$ and PRRFES is right-consistent) and the number of exploration rounds is thus bounded: $K^m_{l+1} < 2^{2^l+1}$. All further steps are similar to (Drutsa, 2017b, Th.5): $\sum_{k=1}^{K^m_l}(v^m - p^m_{l,k}) < 2$; for each phase l, we get that $R^m_l \leq rv^m + 4$; and the number of phases $L^m \leq \log_2\log_2 I^m + 1$. The full proof is in Appendix A.2.1 of Supplementary Materials. □

4.2. Upper Bound for Deviation Regret

Prop. 1 provides us with the tool that locates the valuation v^m of a bidder $m \in \mathbb{M}$ at least in the segment $[u^m_i, w^m_i] := [q^m_{l^m_i}, q^m_{l^m_i} + 2\epsilon_{l^m_i - 1}]$ right after a period $i-1$ (see the proof [sketch] of Lemma 2), when $r \geq r_{\gamma_m}$. This means: if, after playing a period $i-1$, the upper bound w^m_i of the valuation of a bidder $m \in \mathbb{M}$ is lower that the lower bound $u^{\hat{m}}_i$ of the valuation of another bidder $\hat{m} \in \mathbb{M}^{-m}$, i.e., $w^m_i < u^{\hat{m}}_i$, then the bidder m does definitely have non-maximal valuation (i.e., $v^m < \overline{v}$) and needs not to be suspected in the period i and subsequent ones. Hence, based on this observation, one can derive the following stopping rule. For given parameters r and $g(\cdot)$ of the PRRFES algorithm $\mathcal{A}_1$, any state $\mathfrak{n} \in \mathfrak{X}(\mathcal{A}_1)$ of the algorithm can be mapped to the current phase $l(\mathfrak{n})$ and the last accepted price $q(\mathfrak{n})$ before the phase $l(\mathfrak{n})$. Thus, we define the stopping rule by

$$\mathrm{sr}_{\mathcal{A}_1}(m, \{\mathfrak{n}^m\}_{m=1}^M) := \rho(m, \{l(\mathfrak{n}^m)\}_{m=1}^M, \{q(\mathfrak{n}^m)\}_{m=1}^M),$$

where

$$\rho(m, \mathbf{l}, \mathbf{q}) := \exists \hat{m} \in \mathbb{M}^{-m} : q^m + 2\epsilon_{l^m - 1} < q^{\hat{m}} \qquad (8)$$

for any $\mathbf{l} \in \mathbb{Z}^M_+$ and any $\mathbf{q} \in \mathbb{R}^M_+$. The div-transformation $\mathrm{div}_M(\langle\mathcal{A}_1\rangle, \mathrm{sr}_{\mathcal{A}_1})$ of the PRRFES algorithm $\mathcal{A}_1$ with the stopping rule $\mathrm{sr}_{\mathcal{A}_1}$ defined in Eq. (8) is referred to as *the dividing Penalized Reject-Revising Fast Exploiting Search* (*divPRRFES*). The pseudo-code of divPRRFES is presented in Appendix B.2 of Supplementary Materials.

Lemma 3. *Let $\gamma_0 \in (0,1)$, the discounts $\boldsymbol{\gamma} \in (0,\gamma_0]^M$, and the seller use the divPRRFES pricing algorithm $\mathrm{div}_M(\langle\mathcal{A}_1\rangle, \mathrm{sr}_{\mathcal{A}_1})$ with the number of penalization rounds $r \geq r_{\gamma_0}$, with the exploitation rate $g(l) = 2^{2^l}, l \in \mathbb{Z}_+$, and with the stopping rule $\mathrm{sr}_{\mathcal{A}_1}$ defined in Eq. (8). Then, for a bidder $m \in \mathbb{M}$ with non-maximal valuation, i.e., $v^m < \overline{v}$, his subhorizon I^m is bounded:*

$$I^m \leq \frac{24}{\overline{v} - v^m} + r\big(1 + \log_2\log_2 \frac{4}{\overline{v} - v^m}\big) < \frac{24 + 5r}{\overline{v} - v^m}. \quad (9)$$

Proof sketch. Let $\overline{m}$ be a buyer with the maximal valuation $\overline{v}$. Note that, in any period $j = 1, \ldots, I^m$, the location intervals $[q^m_{l^m_j}, q^m_{l^m_j} + 2\epsilon_{l^m_j - 1}]$ and $[q^{\overline{m}}_{l^{\overline{m}}_j}, q^{\overline{m}}_{l^{\overline{m}}_j} + 2\epsilon_{l^{\overline{m}}_j - 1}]$ must intersect (otherwise, the stopping rule $\mathrm{sr}_{\mathcal{A}_1}$ has eliminated the buyer m before the period j, and, hence, $j > I^m$). In particular, in the period I^m,

$$\epsilon_{L(m',m)} \geq \frac{\overline{v} - v^m}{4}$$

holds for either $m' = m$ or (not exclusively) $m' = \overline{m}$, where $L(m', m) := l^{m'}_{I^m}$. From the definition of the iteration parameter ϵ_l, i.e. $\log_2 \epsilon_l = -2^l$, one can obtain the bound on one of the phases:

$$\min\{L(m,m), L(\overline{m}, m)\} \leq \log_2\log_2 \frac{4}{\overline{v} - v^m}. \qquad (10)$$

To bound the subhorizon I^m, decompose it into the numbers of exploration, reject-penalization, and exploitation rounds in each phase $l = 0, \ldots, L(m', m)$ played by a buyer $m' \in \{m, \overline{m}\}$. Applying techniques similar to the ones used in the proof of Lemma 2 (in particular, the bound on the number of exploration rounds: $K^{m'}_l \leq 2 \cdot 2^{2^{l-1}}$), we get:

$$I^m \leq (L(m', m) + 1)r + 2 \cdot 3 \cdot 2^{2^{L(m',m)}} \qquad (11)$$

for $m' \in \{m, \overline{m}\}$. This combined with the previous inequality implies Eq. (9). The full proof is in App. A.2.2. □

This lemma implies the upper bound for the deviation part of the strategic regret of the divPRRFES pricing algorithm $\mathcal{A} = \mathrm{div}_M(\langle\mathcal{A}_1\rangle, \mathrm{sr}_{\mathcal{A}_1})$ against the strategic buyers $\mathbb{M}$:

$$\mathrm{SReg}^{\mathrm{dev}}(T, \mathcal{A}, \mathbf{v}, \boldsymbol{\gamma}, \boldsymbol{\beta}) = \sum_{m=1}^M I^m(\overline{v} - v^m) \leq (24 + 5r)(M - 1).$$

Let us denote by $\overline{\mathbb{M}} := \{m \in \mathbb{M} \mid v^m = \overline{v}\}$ the set of bidders with the maximal valuation and by $\overline{\overline{v}} := \max_{m \in \mathbb{M}\setminus\overline{\mathbb{M}}} v^m$ the highest valuation among non-maximal ones. Thus, we showed that learning of the max-valuation bidders $\overline{\mathbb{M}}$ converges with the rate inversely proportional to $\overline{v} - \overline{\overline{v}}$ (i.e., after the period $\lceil(24 + 5r)/(\overline{v} - \overline{\overline{v}})\rceil$ the set of suspected bidders is always $\mathbb{S}_i = \overline{\mathbb{M}}$) and this learning contributes a constant (w.r.t. the horizon T) to the strategic regret. Finally, Lemma 1, 2, and 3 trivially imply (see Appendix A.2.3) the following theorem.

Theorem 1. *Let $\gamma_0 \in (0,1)$, $\mathcal{A}_1$ be the PRRFES algorithm with $r \geq r_{\gamma_0}$ from Eq. (6) and the exploitation rate $g(l) = 2^{2^l}, l \in \mathbb{Z}_+$, and $\mathrm{sr}_{\mathcal{A}_1}$ be the stopping rule defined in Eq.(8). Then, for $T \geq 2$, the strategic regret of the divPRRFES pricing algorithm $\mathcal{A} = \mathrm{div}_M(\langle\mathcal{A}_1\rangle, \mathrm{sr}_{\mathcal{A}_1})$ against the buyers $\mathbb{M}$ is upper bounded:*

$$\begin{aligned}\mathrm{SReg}(T, \mathcal{A}, \mathbf{v}, \boldsymbol{\gamma}, \boldsymbol{\beta}) \leq & M(r\overline{v} + 4)(\log_2\log_2 T + 2) + \\ & + (24 + 5r)(M - 1) \quad \forall \boldsymbol{\gamma} \in (0, \gamma_0]^M\ \forall \mathbf{v} \in [0,1]^M\ \forall \boldsymbol{\beta}.\end{aligned} \qquad (12)$$

5. Discussion and Extensions of the Result

Other auction formats. The techniques and algorithms developed in our work can be applied in repeated auctions where another format of selling a good in a round is used. Namely, our results hold in our repeated setting with an auction format (within rounds) that satisfies the following: (a) personal reserve prices are allowed; and (b) if a buyer m is only one non-eliminated participant in a round t, then his bidding mechanism allows him to choose between getting the good for the reserve price p_t^m and rejecting it. This holds e.g. for first(/third/..)-price auctions, for PPA with multiple bidders, etc.

Regret dependence on M. The upper bound of the divPRRFES regret in Eq. (12) linearly depends on M. We believe that it is not an artifact of our analysis tools, but a payment for the div-transformation. Consider the case in which all bidders have the same valuation, i.e., all their valuations are $\overline{v}$. Each bidder will be always suspected by divPRRFES (i.e., be in $\mathbb{S}_i \ \forall i$). Hence, divPRRFES will just learn the valuation $\overline{v}$ for each of M bidders independently and, thus, M times slower; i.e., it is natural that the regret of divPRRFES is M times larger than the regret of PRRFES against a single buyer. However, there might exist an algorithm that do not suffer from dividing structure in this way. So, existence of an algorithm with a more favorable regret dependence on M is an open research question.

Improvements of divPRRFES. For practical use, there are several places where divPRRFES can be improved. For instance, (a) the penalization parameter r can be made adaptive to take into account the rounds in which a buyer is eliminated (i.e., reduce the number of penalizations by the number of rivals currently suspected by the seller); (b) or the stopping rule $\mathrm{sr}_{\mathcal{A}_1}$ can faster eliminate bidders, since the lower bound u_i^m can be updated each time the buyer m accepts an exploration price $p_{l,k}^m$. Despite these improvements would require some additional pages in our proofs, they do not improve the asymptotic bound of $O(\log\log T)$. The constants in the regret bound Eq. (12) can be optimized in a way similar to the one applied by Drutsa (2017a) for repeated posted-price auctions.

Lower bound and optimality. For the case $M = 1$, there does exist the lower bound: the strategic regret of any pricing algorithm is $\Omega(\log\log T)$ (Mohri & Munoz, 2014). Hence, our upper bound for the algorithm divPRRFES is optimal in the general case of any number of bidders. Nonetheless, structure of the game with non-single buyer ($M \geq 2$) is much more complicated, since a buyer has to act in the presence of rivals and under uncertainty about the future. This is an additional opportunity that can be exploited by a pricing algorithm. Thus, the validity of the lower bound $\Omega(\log\log T)$ for $M \geq 2$ is an open research question.

Horizon independence. The algorithm divPRRFES is horizon-independent since it is based on the horizon-independent PRRFES $\mathcal{A}_1$, which induces the subalgorithm $\langle\mathcal{A}_1\rangle$ and the stopping rule $\mathrm{sr}_{\mathcal{A}_1}$. Hence, the seller is not required to know in advance the number of rounds T of the game, when she applies divPRRFES.

6. Conclusions

We studied the scenario of repeated second-price auctions with reserve pricing where a seller interacts with multiple strategic buyers. Each buyer participates in each round of the game, holds a fixed private valuation for a good, and seeks to maximize his expected future discounted surplus; while the seller seeks a no-regret online learning algorithm to set reserve prices for worst-case valuations. First, we proposed the so-called dividing transformation that upgrades an algorithm designed for the setup with a single buyer to the multi-buyer case. Second, the transformation allowed us to obtain a novel horizon-independent algorithm that can be applied against strategic buyers with regret upper bound of $O(\log\log T)$. Finally, we introduced non-trivial techniques such as (a) the method to locate the valuation of a strategic buyer in a played round under buyer uncertainty about the future; (b) the decomposition of strategic regret into the individual and deviation parts; and (c) the approach to learn the highest-valuation bidder with deviation regret of $O(1)$.

ACKNOWLEDGMENTS

I would like to thank Sergei Izmalkov who inspired me to conduct this study.

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

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

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

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

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

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

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

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

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.

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

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

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

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

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

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. Optimal non-parametric learning in repeated contextual auctions with strategic buyer. 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.

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

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.

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.

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

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

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

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

Leme, R. P. and Schneider, J. Contextual search via intrinsic volumes. *arXiv preprint arXiv:1804.03195*, 2018.

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

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

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.

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

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.

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

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

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

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.