Learning to Bid in Repeated First-Price Auctions with Budgets

Qian Wang [*1] **Zongjun Yang** [*2] **Xiaotie Deng** [1 3] **Yuqing Kong** [1]

Abstract

Budget management strategies in repeated auctions have received growing attention in online advertising markets. However, previous work on budget management in online bidding mainly focused on second-price auctions. The rapid shift from second-price auctions to first-price auctions for online ads in recent years has motivated the challenging question of how to bid in repeated first-price auctions while controlling budgets. In this work, we study the problem of learning in repeated first-price auctions with budgets. We design a dual-based algorithm that can achieve a near-optimal $\widetilde{O}(\sqrt{T})$ regret with full information feedback where the maximum competing bid is always revealed after each auction. We further consider the setting with one-sided information feedback where only the winning bid is revealed after each auction. We show that our modified algorithm can still achieve an $\widetilde{O}(\sqrt{T})$ regret with mild assumptions on the bidder's value distribution. Finally, we complement the theoretical results with numerical experiments to confirm the effectiveness of our budget management policy.

1. Introduction

Recent years have witnessed the explosive growth of the online advertising market. It is estimated that worldwide online advertising spending will reach 681 billion dollars in 2023, accounting for nearly 70% of the entire advertising market spending (eMarketer, 2022). In practice, a huge amount of online ads are sold via real-time auctions implemented on advertising platforms and advertisers participate in such repeated online auctions to purchase advertising opportunities. An advertiser typically aims to maximize her cumulative payoffs during a specific time horizon (e.g., a day or a week) subject to a budget constraint, which reflects her monetary limit throughout this period. The presence of budgets constitutes an important operational challenge for online bidding algorithm design since there will be considerable financial losses whether depleting the budget too early or reaching the end of the period with unused funds. Therefore, budget management is a fundamental issue for designing practical online bidding algorithms.

There has been a flourishing line of literature on budget management strategies in repeated second-price auctions (Balseiro & Gur, 2019; Balseiro et al., 2022b; Chen et al., 2022). However, a major industry-wide shift has occurred recently towards using first-price auctions as the preferred auction format of selling digital ads (Despotakis et al., 2021), as opposed to the earlier prevalent practice of using second-price auctions (Lucking-Reiley, 2000; Klemperer, 2004; Lucking-Reiley et al., 2007). Google Ad Exchange, the largest online auction platform, announced its shift to the first-price auction in September 2019 (Bigler, 2019). On the one hand, the difference in the nature of first-price auctions and second-price auctions implies that the algorithms proposed by the above work may not directly apply to the first-price setting. On the other hand, most previous work on repeated first-price auctions only considered bidding without budgets (Balseiro et al., 2019; Han et al., 2020b;a; Badanidiyuru et al., 2021; Zhang et al., 2022). The shift thus leads to a pressing question of how should an advertiser bid in repeated first-price auctions to maximize the cumulative payoffs while controlling the expenditures.

In this paper, we study the design of online bidding algorithms in repeated first-price auctions with budgets. We focus on a *stochastic* setting where both the bidder's values and the maximum competing bids are *i.i.d.* sampled over auctions. The goal is to optimize the bidder's expected cumulative rewards while keeping her budget constraint satisfied for any realization of values and competing bids. We provide online bidding algorithms for the bidder in two different feedback models: (1) the *full information feedback*, where the maximum competing bid is always revealed after each auction; (2) the *one-sided information feedback*, where

[*]Equal contribution [1]Center on Frontiers of Computing Studies, School of Computer Science, Peking University, Beijing, China [2]School of Electronics Engineering and Computer Science, Peking University, Beijing, China [3]Center for Multi-Agent Research, Institute for AI, Peking University, Beijing, China. Correspondence to: Qian Wang <charlie@pku.edu.cn>, Zongjun Yang <allenyzj@stu.pku.edu.cn>, Xiaotie Deng <xiaotie@pku.edu.cn>, Yuqing Kong <yuqing.kong@pku.edu.cn>.

Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023. Copyright 2023 by the author(s).

only the winning bid is revealed after each auction.

Our Results. Our main contribution is to propose two near-optimal bidding algorithms for each of the two feedback models. For the full information feedback, Algorithm 1 can achieve an $\widetilde{O}(\sqrt{T})$ regret where T is the total number of auctions (Theorem 3.2). For the the one-sided information feedback, Algorithm 2 can achieve an $\widetilde{O}(\sqrt{T})$ regret under mild assumptions on value distributions (Theorem 3.7).

Our algorithms follow a primal-dual framework and update a dual variable via online gradient descent to adjust the rate at which the bidder depletes her budget. The framework is similar to those used by previous work on second-price auctions (Balseiro & Gur, 2019; Balseiro et al., 2022b; Feng et al., 2022; Chen et al., 2022). However, its application to repeated first-price auctions presents new challenges:

1. First, due to the non-truthful nature of first-price auctions, we cannot compute the bid that maximizes the cost-adjusted reward without knowing the highest competing bid. One can only expect to maximize the objective in expectation by learning the hidden distribution of the maximum competing bids. However, estimates using historical samples may not be sufficiently accurate. Even worse, with one-sided feedback, the bidder's observations are actually biased.

2. Second, the dynamic update of the dual variable implies that the cost-adjusted reward function differs in different rounds. This causes failure in previous bandit algorithms like Balseiro et al. (2019) and Han et al. (2020b), as future rounds may suffer from exploitation when objectives are misaligned.

This work overcomes the two challenges and provides theoretical performance guarantees for our proposed algorithms. We start with the full information feedback and design an algorithm (Algorithm 1), which sketches the high-level combination of online optimization methods and distribution estimation techniques. We then refine the algorithm for the one-sided information feedback, where we introduced value shading (i.e., scaling down the value by a factor) to align the objectives of different rounds so as to balance exploration and exploitation. We maintain a high-reward bid set for each shaded value and leverage the graph-feedback and partial-order properties in first-price auctions, which in essence follows the approach developed in Han et al. (2020b). As the bidder's observations are biased in this feedback model, the estimation errors are bounded via a martingale argument. The sum of estimation errors is shown to be upper bounded by $\widetilde{O}(\sqrt{T})$ with an assumption on the bidder's value distribution. In the experimental part, we demonstrate that our algorithms outperform those without budget management under various distributions in terms of the long-run average performance.

1.1. Related Work

Learning in repeated auctions with constraints has been extensively studied in literature but most studies focused on only second-price auctions. Balseiro & Gur (2019) proposed an optimal online bidding algorithm known as adaptive pacing in repeated second-price auctions with budget constraints. Balseiro et al. (2022b) extended the above work by relaxing some model assumptions and using a more general dual approximation scheme. Feng et al. (2022); Golrezaei et al. (2021) considered repeated second-price auctions with both budget and return-on-spend (RoS) constraints. Chen et al. (2022) studied another important class of budget management strategies, called throttling, and proposed a near-optimal throttling algorithm. Our work considers budget management strategies in repeated first-price auctions instead but we use some dual-based techniques similar to those in the above papers.

For repeated first-price auctions, most previous work only considered bidding without constraints. Balseiro et al. (2019) first considered learning in repeated first-price auctions with binary feedback. They adopted a cross-learning approach to achieve an $\widetilde{O}(T^{2/3})$ regret and showed that the lower bound on regret is $\Omega(T^{2/3})$. Han et al. (2020b) considered repeated first-price auctions with one-sided feedback. They leveraged the graph-feedback and partial-order properties in first-price auctions to achieve an $\widetilde{O}(\sqrt{T})$ regret and proved that the lower bound on regret is $\Omega(\sqrt{T})$ even under full feedback. Han et al. (2020a) considered the setting without the *i.i.d.* assumption of the maximum competing bids but with a Lipschitz assumption on competing strategies. Badanidiyuru et al. (2021) studied contextual first-price auctions where the values and the maximum competing bids depend on public contexts. Zhang et al. (2022) studied the setting where the bidder has access to some hint relevant to the maximum competing bid. The main difference between the above papers and ours is that the bidder has a budget constraint in our model.

Technically, our setting is closely related to the *Bandits with Knapsacks* (BwK) literature (Badanidiyuru et al., 2018; Agrawal & Devanur, 2019; Castiglioni et al., 2022; Immorlica et al., 2022), among which Castiglioni et al. (2022) considered first-prices auctions with budgets, as an application of the BwK framework. Given a finite value set of size n_v and a finite bid set of size n_b, their algorithm can achieve an $O(\sqrt{Tn_v \log(n_b)})$ regret in various feedback models by instantiating appropriate regret minimizers. Hence, with a careful discretization, their result immediately implies a sublinear regret bound for the setting with continuous valuations and bids. However, taking into account the dis-

cretization error of order $O(T/n_v + T/n_b)$, the best regret guarantee they can obtain is $\widetilde{O}(T^{2/3})$, while we can achieve a near-optimal $\widetilde{O}(\sqrt{T})$ regret for the continuous setting.

Ai et al. (2022) studied a variant problem of repeated first-price auctions with budgets. Their model additionally involves a discount factor $\gamma < 1$ in the objective function and the "optimal" strategy is defined with respect to this variant problem. Their algorithm will abort if $\gamma = 1$ so our results are not directly comparable with theirs.

Balseiro et al. (2022a) studied the equilibrium bidding strategies for first-price auctions with budgets. Their value-pacing-based strategies are similar to our second algorithm, but note that we are investigating a dynamic setting from the view of a single budget-constrained bidder. The equilibrium characterization of the first-price market cannot provide regret guarantees for dynamic bidding, especially considering the learning process with respect to different feedback models.

2. Model and Benchmark

In this work, we consider the problem of online learning in the first-price auction market. We focus on a single bidder in a large population of bidders during a time horizon T.

In each round $t = 1, \ldots, T$, there is an available ad slot auctioned by the seller (e.g. an advertising platform). The bidder receives a private value $v_t \in [0, \bar{v}]$, and then submits a bid $b_t \in \mathbb{R}_+$ based on v_t and all historical observations available to her. We denote the maximum bid of all other bidders by $d_t \in \mathbb{R}_+$. The auction outcome depends on the comparison between b_t and d_t. Let $x_t \coloneqq \mathbf{1}\{b_t \geq d_t\}$ be the binary variable indicating whether the bidder wins the ad slot at round t. Here we assume that ties are broken in favor of the bidder we concern to simplify exposition. We note that this choice is arbitrary and by no means a limitation of our approach. Let $r_t \coloneqq x_t(v_t - b_t)$ be her reward and let $c_t \coloneqq x_t b_t$ be the corresponding cost for a first-price auction. As usual, we use the bold symbol $\boldsymbol{v}$ without subscript t to denote the vector $(v_1, \ldots, v_T)$; the same goes for other variables in the present paper.

The bidder has a budget B that limits the payments she can make over T rounds of auctions, and her maximum expenditure rate is denoted by $\rho \coloneqq B/T$. We assume that $\rho \in (0, \bar{v}]$; otherwise, it becomes a problem without constraints as the bidder would never deplete her budget.

We consider a *stochastic* setting where v_t is *i.i.d.* sampled from a distribution F and d_t is *i.i.d.* sampled from a distribution G. The latter assumption follows from the standard mean-field approximation (Iyer et al., 2014; Balseiro et al., 2015) and is a common practice in literature. The main rationale behind this assumption is that when the number of other bidders is large, on average their valuations and bidding strategies are static over time. Note that both F and G are unknown to the bidder.

Information structure. In repeated first-price auctions, the bidder can receive different feedback after each round depending on the information released by the seller. In particular, the ability of this bidder to observe the maximum competing bid d_t varies to the information structure. In this paper, we investigate two different information structures:

1. *Full information feedback.* The bidder can observe the maximum competing bid d_t at the end of each round t. This information structure makes sense in many current online auction platforms. For example, in the Google Ad Exchange, at the end of an auction, bidders will receive back the minimum value they would have had to bid to win the auction, whether they lose or win (Google Ad Exchange, 2022).
2. *One-sided information feedback.* The bidder can observe the maximum competing bid d_t only if she loses the auction. Thus, the feedback available to her includes $\mathbf{1}\{b_t \geq d_t\}$ and $d_t\mathbf{1}\{b_t < d_t\}$. This can be viewed as an informational version of the *winner's curse* (Capen et al., 1971) where the winner learns less information. And this is a common feedback model in previous studies on repeated first-price auctions (Esponda, 2008; Han et al., 2020b; Ai et al., 2022). Compared to the full information feedback, the one-sided information feedback is more complicated to deal with as the bidder can observe less information.

We denote the historical observations available to the bidder before submitting a bid in round t by $\mathcal{H}_t$. For the full information feedback, we define

$$\mathcal{H}_t^F \coloneqq (v_s, x_s, d_s)_{s=1}^{t-1}.$$

At the end of each round s, the bidder can append a tuple (v_s, x_s, d_s) to the available history. For the one-sided information feedback, we define

$$\mathcal{H}_t^O \coloneqq (v_s, x_s, (1 - x_s)d_s)_{s=1}^{t-1}.$$

At the end of each round s, the bidder knows her value and whether she wins, but the winner can only observe $(1 - x_s)d_s = 0$.

Bidding strategy and regret. A bidding strategy maps $(\mathcal{H}_t, v_t)$ to a (possibly random) bid b_t for each t. We say π is *budget feasible* if it generates expenditures that are constrained by the budget for any realizations of values and maximum competing bids, i.e. $\forall \boldsymbol{v}, \boldsymbol{d}$,

$$\sum_{t=1}^{T} c_t^\pi = \sum_{t=1}^{T} \mathbf{1}\{b_t^\pi \geq d_t\} b_t^\pi \leq B = \rho T.$$

We denote by Π_0 the set of all budget feasible strategies. For a strategy $\pi \in \Pi_0$, we denote by $R(\pi)$ the performance of π, defined as follows:

$$\begin{aligned} R(\pi) =& \mathbb{E}^{\pi}_{\boldsymbol{v},\boldsymbol{d}} \left[\sum_{t=1}^{T} r_t^{\pi} \right] \\ =& \mathbb{E}^{\pi}_{\boldsymbol{v},\boldsymbol{d}} \left[\sum_{t=1}^{T} \mathbf{1}\left\{ b_t^{\pi} \geq d_t \right\} (v_t - b_t^{\pi}) \right], \end{aligned}$$

where the expectation is taken with respect to the values, the maximum competing bids and any possible randomness embedded in the strategy. The bidder's optimization problem can be written as

$$\begin{aligned} \max_{\pi} \ & \mathbb{E}^{\pi}_{\boldsymbol{v},\boldsymbol{d}} \left[\sum_{t=1}^{T} \mathbf{1}\left\{ b_t^{\pi} \geq d_t \right\} (v_t - b_t^{\pi}) \right] \\ \text{s.t.} \ & \sum_{t=1}^{T} \mathbf{1}\left\{ b_t^{\pi} \geq d_t \right\} b_t^{\pi} \leq \rho T, \ \forall \boldsymbol{v}, \boldsymbol{d}. \end{aligned} \tag{1}$$

The regret of the bidder is defined to be the difference in the expected cumulative rewards of the bidder's strategy and the optimal budget feasible bidding strategy, which has the perfect knowledge of F and G:

$$Reg(\pi) = \max_{\pi' \in \Pi_0} R(\pi') - R(\pi).$$

3. Bidding Algorithms and Analysis

In this section, we first design and analyze an algorithm for Problem (1) with full information feedback, which reveals our high-level idea on budget management in repeated first-price auctions. We prove that the algorithm can achieve an $\widetilde{O}(\sqrt{T})$ regret. Then we modify our algorithm to accommodate the setting with only one-sided information feedback by using value shading and leveraging a special partial order property possessed by first-price auctions. The modified algorithm can still achieve an $\widetilde{O}(\sqrt{T})$ regret under mild assumptions. All omitted proofs in this section can be found in the appendix.

3.1. Full Information Feedback

Our bidding algorithm for full information feedback is depicted in Algorithm 1. The bidder first conducts a one-round exploration to make an appropriate initialization (Line 2). After observing the value v_t in each round $t = 2, \ldots, T$, the bidder constructs estimates of empirical rewards and costs based on all historical observations and submits a bid that maximizes a cost-adjusted reward (Lines 4 to 6). Then the bidder updates λ_t using the empirical cost of the submitted bid $\widetilde{c}_t(b)$ and her average budget ρ (Line 7). The variable λ_t plays a key role in adjusting the pace at which the bidder depletes her budget. By the choice of b_t, we must have $b_t \leq v_t/(1+\lambda_t)$. If the bidder bids too high in past rounds, λ_t tends to be larger, thereby controlling the bids in future rounds.

Algorithm 1 Bidding Algorithm for First-Price Auctions with Budgets under Full Information Feedback

1: **Input:** Time horizon T; budget $B = \rho T$; update step $\epsilon > 0$; failure probability $\delta > 0$.
2: **Initialization:** The bidder bids $b_1 = 0$ and set $B_2 = B, \lambda_2 = 0$.
3: **for** $t \in \{2, \cdots, T\}$ **do**
4: The bidder receives the value $v_t \in [0, \bar{v}]$.
5: The bidder estimates the rewards and costs:

$$\widetilde{r}_t(v_t, b) = \frac{1}{t-1} \sum_{s=1}^{t-1} \mathbf{1}\left\{ b \geq d_s \right\} (v_t - b), \tag{2}$$

$$\widetilde{c}_t(b) = \frac{1}{t-1} \sum_{s=1}^{t-1} \mathbf{1}\left\{ b \geq d_s \right\} b. \tag{3}$$

6: The bidder submits a bid:

$$b_t \in \arg\max_{b} \left(\widetilde{r}_t(v_t, b) - \lambda_t \widetilde{c}_t(b) \right). \tag{4}$$

(Taking the smallest if there are ties.)
7: The bidder updates the parameter

$$\lambda_{t+1} = \mathrm{Proj}_{\lambda > 0} \left(\lambda_t - \epsilon \left(\rho - \widetilde{c}_t(b_t) \right) \right). \tag{5}$$

8: The bidder observes the maximum competing bid d_t.
9: The bidder update the remaining budget

$$B_{t+1} = B_t - c_t. \tag{6}$$

10: **if** $B_{t+1} < \bar{v}$ **then**
11: **break**
12: **end if**
13: **end for**

To provide more intuition on the choice of b_t and the update procedure of λ_t, we consider an alternative optimization problem with a soft budget constraint:

$$\begin{aligned} \max_{\pi} \ & \mathbb{E}^{\pi}_{\boldsymbol{v},\boldsymbol{d}} \left[\sum_{t=1}^{T} \mathbf{1}\left\{ b_t^{\pi} \geq d_t \right\} (v_t - b_t^{\pi}) \right] \\ \text{s.t.} \ & \mathbb{E}^{\pi}_{\boldsymbol{v},\boldsymbol{d}} \left[\sum_{t=1}^{T} \mathbf{1}\left\{ b_t^{\pi} \geq d_t \right\} b_t^{\pi} \right] \leq \rho T. \end{aligned} \tag{7}$$

The Lagrangian dual objective of Problem (7) is

$$\mathbb{E}^{\pi}_{\boldsymbol{v},\boldsymbol{d}}\left[\sum_{t=1}^{T}\left(\mathbf{1}\left\{b_t^{\pi}\geq d_t\right\}\left(v_t-(1+\lambda)b_t^{\pi}\right)+\lambda\rho\right)\right]$$
$$=\sum_{t=1}^{T}\left(\mathbb{E}^{\pi}_{\mathcal{H}_t,v_t}\left[\left(v_t-(1+\lambda)b_t^{\pi}\right)G\left(b_t^{\pi}\right)\right]+\lambda\rho\right),$$

where the equality holds since b_t^{π} is independent of d_t as well as other future values and maximum competing bids. For a fixed λ, the dual objective is maximized by bidding $b_t\in\arg\max_b(v_t-(1+\lambda)b)G(b)$, which is irrelevant to the historical observations $\mathcal{H}_t$.

We denote by Π_1 the set of all strategies that satisfy the soft budget constraint in Problem (7). By weak duality, we have

$$\max_{\pi\in\Pi_1}R(\pi)$$
$$\leq\min_{\lambda\geq 0}T\left(\mathbb{E}_v\left[\max_b\left(v-(1+\lambda)b\right)G\left(b\right)\right]+\lambda\rho\right). \quad (8)$$

Our algorithm adopts an online gradient descent scheme to approximate the right hand side of (8), which is also an upper bound on the optimal value of Problem (1) since $\Pi_0\subseteq\Pi_1$. A crucial challenge here is that the bidder does not know the prior distribution G so she cannot calculate the exact maximum point of $(v-(1+\lambda)b)G(b)$. To deal with this issue, we adopt the distribution estimation method and use $\widetilde{r}_t(v_t,b),\widetilde{c}_t(b)$ in place of $(v-b)G(b),bG(b)$. As t grows, the estimates become more accurate. The following lemma shows that $\forall v_t,b$, $\widetilde{r}_t(v_t,b)$ and $\widetilde{c}_t(b)$ are good estimates of $r(v_t,b):=(v_t-b)G(b)$ and $c(b):=bG(b)$ respectively.

Lemma 3.1. *Under Algorithm 1, with probability at least $1-\delta$, we have for all $t\geq 2$ and $b\leq\bar{v}$,*

$$|\widetilde{r}_t(v_t,b)-r(v_t,b)|\leq\bar{v}\cdot\sqrt{\frac{\ln\left(2T/\delta\right)}{2(t-1)}}, \quad (9)$$

$$|\widetilde{c}_t(b)-c(b)|\leq\bar{v}\cdot\sqrt{\frac{\ln\left(2T/\delta\right)}{2(t-1)}}. \quad (10)$$

Theorem 3.2. *For repeated first-price auctions with budget constraints and full information feedback, Algorithm 1 can achieve*

$$Reg(\pi)=O\left(\sqrt{T\ln T}\right).$$

In the proof of Theorem 3.2, we first we perform a standard analysis of the online gradient descent method to show the sequence of λ_t is not much worse than a hindsight λ with respect to gain function $h_t(\lambda_t)=\lambda_t\left(\widetilde{c}_t(b)-\rho\right)$. This, together with Lemma 3.1 implies that with high probability, the bid b_t chosen in each round is close to the bid chosen by the best strategy for Problem (7), i.e., $\arg\max_b(v_t-(1+\lambda^*)b)G(b)$ where λ^* is the optimal dual variable. Finally, the proof is concluded by showing that the time at which the budget is depleted under Algorithm 1 is close to T.

Lower bound. As previous work has proved, the lower bound on regret for this problem is $\Omega(\sqrt{T})$ even without constraints (i.e., $\rho=\bar{v}$).

Lemma 3.3 (Han et al. (2020b))**.** *For repeated first-price auctions and full information feedback, there exists a positive constant $C>0$ independent of T such that*

$$\inf_{\pi}\sup_{F,G}Reg(\pi)\geq C\sqrt{T}. \quad (11)$$

This previous result implies that our algorithm can achieve a near-optimal learning performance in first-price auctions with budget constraints.

Discretization. In Algorithm 1, Line 5 requires estimating rewards and costs for all possible bids but the bid space might be continuous. In practice, we can resolve this issue by a simple discretization, which will cause little performance degradation. Let $\mathcal{B}=\{b^1,\cdots,b^K\}$ with $b^k=(k-1)/K\cdot\bar{v}$. We then change Line 6 to that the bidder submits a bid

$$b_t\in\arg\max_{b\in\mathcal{B}}\left(\widetilde{r}_t(v_t,b)-\lambda_t\widetilde{c}_t(b)\right). \quad (12)$$

When $K=\Omega(\sqrt{T})$, the discretization error is of order $O(\sqrt{T})$. In the next subsection, we will discretize both values and bids in Algorithm 2 and more formally analyze the additional regret caused by the discretization.

3.2. One-sided Information Feedback

This subsection provides a modified algorithm for the scenario with one-sided information feedback. We show an $\widetilde{O}(\sqrt{T})$ regret can still be achieved with an assumption on the bidder's value distribution.

As the bidder can only observe the highest competing bid d_s after losing at round s under one-sided information feedback, she can no longer estimate the expected rewards and costs in each round using all past rounds as in Algorithm 1. Specifically, given value v_t and bid b, if $b<b_s$ and $b_s\geq d_s$, she cannot determine $\mathbf{1}\{b\geq d_s\}$, so that she cannot calculate $\widetilde{r}_t(v_t,b),\widetilde{c}_t(b)$ as in (2), (3). Therefore, we need new estimators for the expected rewards and costs.

For this purpose, we first discretize the value space into a set of size M, $\mathcal{V}=[v^1,\cdots,v^M]$ with $v^m=(m-1)/M\cdot\bar{v}$, and the bid space into a set of size K, $\mathcal{B}=\{b^1,\cdots,b^K\}$ with $b^k=(k-1)/K\cdot\bar{v}$. Then, we denote by n_t^k the number of observed bids lower than b^k before round t:

$$n_t^k:=\sum_{s=1}^{t-1}\mathbf{1}\left\{b_s\leq b^k\right\}. \quad (13)$$

Given v^m and b^k, we define two estimators under one-sided

information feedback as

$$\widetilde{r}_t(v^m, b^k) = \frac{1}{n_t^k} \sum_{s=1}^{t-1} \mathbf{1}\left\{b_s \le b^k\right\} \mathbf{1}\left\{b^k \ge d_s\right\} \left(v^m - b^k\right), \tag{14}$$

$$\widetilde{c}_t(b^k) = \frac{1}{n_t^k} \sum_{s=1}^{t-1} \mathbf{1}\left\{b_s \le b^k\right\} \mathbf{1}\left\{b^k \ge d_s\right\} b^k. \tag{15}$$

Note that the equations (14) and (15) are measurable with respect to the available history $\mathcal{H}_t^O$. If $b_s \ge d_s$, we have $\mathbf{1}\left\{b_s \le b^k\right\} \mathbf{1}\left\{b^k \ge d_s\right\} = \mathbf{1}\left\{b_s \le b^k\right\}$; if $b_s \le d_s$, the bidder can observe the exact d_s to determine $\mathbf{1}\left\{b^k \ge d_s\right\}$.

Since $(d_s)_{s=1}^{t-1}$ are not mutually independent conditioned on $(b_s)_{s=1}^{t-1}$, $\widetilde{r}_t(v^m, b^k)$ and $\widetilde{c}_t(b^k)$ are actually not unbiased estimators. However, we can still prove via a martingale argument that they approximate well to the expected reward $r(v^m, b^k) = (v^m - b^k)G(b^k)$ and the expected cost $c(b^k) = b^k G(b^k)$ with high probability, which is formalized as the following lemma.

Lemma 3.4. *Under Algorithm 2, with probability at least $1-\delta$, we have $\forall t \ge 2, m \in [M], k \in [K]$,*

$$|\widetilde{r}_t(v^m, b^k) - r(v^m, b^k)| \le \bar{v} \cdot \sqrt{\frac{4 \ln T \ln (KT/\delta)}{n_t^k}}, \tag{16}$$

$$|\widetilde{c}_t(b^k) - c(b^k)| \le \bar{v} \cdot \sqrt{\frac{4 \ln T \ln (KT/\delta)}{n_t^k}}. \tag{17}$$

By Lemma 3.4, we know that the more bids that are lower than b^k, the more accurate the estimation of $r(v^m, b^k)$ and $c(b^k)$. However, bidding low in order to benefit future estimates may cause great loss in the current round. The existence of the budget constraint further increases the difficulty of balancing present and future rewards.

We depict our modified algorithm for one-sided information feedback in Algorithm 2. The main difference with Algorithm 1 is that the bidder maintains an active set of high reward bids for each $m \in [M]$. After observing the value v_t in round t, the bidder shades it by $(1+\lambda_t)$ (Line 11) and rounds it to $v^{m(t)} \in \mathcal{V}$. Then the bidder submits a bid among the active bid set $B_t^{m(t)}$ (Line 12). The value-shading step essentially aligns the objectives of different rounds. Observe that

$$\begin{aligned} r(v_t, b^k) - \lambda_t c(b^k) &= \left(v_t - (1+\lambda_t)\, b^k\right) G(b^k) \\ &= (1+\lambda_t) \left(\frac{v_t}{1+\lambda_t} - b^k\right) G(b^k) \\ &= (1+\lambda_t) \cdot r(v_t/(1+\lambda_t), b^k). \end{aligned}$$

Thus, with $v^m \approx v_t/(1+\lambda_t)$, maximizing $r(v_t, b) - \lambda_t c(b)$ is approximately equivalent to maximizing the reward

Algorithm 2 Bidding Algorithm for First-Price Auctions with Budgets under One-Sided Information Feedback

1: **Input:** Time horizon T; budget $B = \rho T$; value set $\mathcal{V} = [v^1, \cdots, v^M]$ with $v^m = (m-1)/M \cdot \bar{v}$; bid set $\mathcal{B} = \{b^1, \cdots, b^K\}$ with $b^k = (k-1)/K \cdot \bar{v}$; update step $\epsilon > 0$; failure probability $\delta \in (0,1)$.
2: **Initialization:** The bidder bids $b_1 = 0$ and set $B_2 = B, \lambda_2 = 0$. Set $\mathcal{B}_0^m \leftarrow \mathcal{B}$ for each $v_m \in \mathcal{V}$.
3: **for** $t \in \{2, \cdots, T\}$ **do**
4: The bidder receives the value $v_t \in [0,1]$;
5: The bidder counts the observations by (13), and estimates the rewards and costs by (14) and (15).
6: **for** $m \in \{1, 2, \cdots, M\}$ **do**
7: The bidder eliminates bids by:

$$\mathcal{B}_{t-1}^m = \left\{ b^k \in \mathcal{B}_{t-1}^m : b^k \ge \max_{s<m} \inf \mathcal{B}_t^s \right\}. \tag{18}$$

8: The bidder computes the confidence bound:

$$w_t^m = \bar{v} \cdot \sqrt{\frac{4 \ln T \log(KT/\delta)}{N_t^m}}, \tag{19}$$

where $N_t^m = \min_{b^k \in \mathcal{B}_{t-1}^m} n_t^k$.
9: The bidder eliminates bids by:

$$\mathcal{B}_t^m \leftarrow \Big\{ b^k \in \mathcal{B}_{t-1}^m : \widetilde{r}_t(v^m, b^k) \ge \max_{b^{k'} \in \mathcal{B}_{t-1}^m} \widetilde{r}_t(v^m, b^{k'}) - 2w_t^m \Big\}. \tag{20}$$

10: **end for**
11: The bidder chooses

$$v^{m(t)} = \max\{u \in \mathcal{V} : u \le v_t/(1+\lambda_t)\}. \tag{21}$$

12: The bidder submits a bid $b_t = \inf \mathcal{B}_t^{m(t)}$;
13: The bidder updates the parameter

$$\lambda_{t+1} = \mathrm{Proj}_{\lambda > 0}\left(\lambda_t - \epsilon\left(\rho - \widetilde{c}_t(b_t)\right)\right). \tag{22}$$

14: The bidder observes x_t and $(1-x_t)d_t$;
15: The bidder update the remaining budget

$$B_{t+1} = B_t - c_t. \tag{23}$$

16: **if** $B_{t+1} < \bar{v}$ **then**
17: **break**
18: **end if**
19: **end for**

$r(v^m, b)$ as if the bidder is participating in a first-price auction without any budget constraint and the value is v^m.

Instead of submitting the bid with the highest estimated reward $\widetilde{r}_t(v^m, b^k)$, the algorithm chooses the smallest bid in $\mathcal{B}_t^m$ in order to make future estimates as accurate as possible. In addition to filtering bids according to the expected rewards and confidence bounds (Line 9), the algorithm also eliminates some small bids from the active sets in Line 7. This elimination increases N_t^m so that the bidder can use smaller confidence bounds to prune the active sets. As long as the best bid for a value v^m remains in the active set $\mathcal{B}_t^m$, the elimination can further control the regret.

To prove that the best bids are not eliminated, we leverage a special partial order property of first price auctions, i.e., $b^*(v) = \arg\max_b (v-b)G(b)$ is non-decreasing in v. In particular, for our scenario with discretization, $\widetilde{b}(v) = \arg\max_{b\in\mathcal{B}}(v-b)G(b)$ is non-decreasing in v. The property guarantees that with high probability, $\widetilde{b}(v^m)$, which approximately maximizes $r(v^m, b)$, is not eliminated from the active set B_{t-1}^m by Line 7. (See Lemma B.3.)

Assumption 3.5. The cumulative probability distribution F is continuous with bounded density function f satisfying $0 < \underline{f} < f(v) < \overline{f} < \infty$ for $v \in [0, \bar{v}]$.

Assumption 3.5 is a technical assumption required by our analysis. We note that the existence of positive bounds on the density function is a common assumption in various learning problems. For example, Balseiro & Gur (2019) took it as one of the sufficient conditions for the strong convexity of the dual objective function.

Lemma 3.6. *Suppose that Assumption 3.5 holds. We have*

$$\mathbb{E}_{\boldsymbol{v},\boldsymbol{d}}\left[\sum_{t=2}^{T}\sqrt{1/N_t^{m(t)}}\right] \leq \widetilde{O}\left(\sqrt{T}\right).$$

For full information feedback, the estimation error in round t is $O(\sqrt{\ln T/(t-1)})$ by Lemma 3.1 so that we are able to control the sum of errors within $\widetilde{O}(\sqrt{T})$. Lemma 3.6 establishes that for one-sided information feedback, we can get a similar result given Assumption 3.5 holds, which constitutes a key part of the proof of the following Theorem 3.7.

Theorem 3.7. *Suppose that Assumption 3.5 holds. For repeated first-price auctions with budget constraints and one-sided information feedback, there exists constants C_1, C_2, C_3, such that Algorithm 2 can achieve*

$$Reg(\pi) \leq C_1\sqrt{T\ln(KT^2)}\ln T + C_2\frac{T}{M} + C_3\frac{T}{K}.$$

Particularly, when choosing $K = M = O(\sqrt{T})$, we obtain that $Reg(\pi) \leq \widetilde{O}(\sqrt{T})$ when T is sufficiently large. We note that to prove Theorem 3.7, Assumption 3.5 is sufficient but may not be necessary. In Section 4, we run numerical experiments with further discussion.

4. Experiments

In this section, we empirically evaluate the reward obtained by our proposed algorithms with both full and partial information feedback, using data generated from various distributions. The primary objective of the numerical experiments is to demonstrate the effectiveness of budget management in different settings. The performance may be further improved by tuning the parameters according to the amount of available budgets. The characterization of such optimal context-dependent parameters is left as an open future problem.

Setup. We consider repeated first-price auctions with $T = 10^6$ rounds, budget amount $B = 10^4$ and upper bound on values $\bar{v} = 1$. We generate the sequence of competing bids by sampling each d_t *i.i.d.* from normal distribution $\mathcal{N}(0.4, 0.1)$. For the sequence of private values, we consider normal distribution, logarithmic normal distribution and uniform distribution respectively. Detailed parameters of the private value distributions can be found in Figure 1.

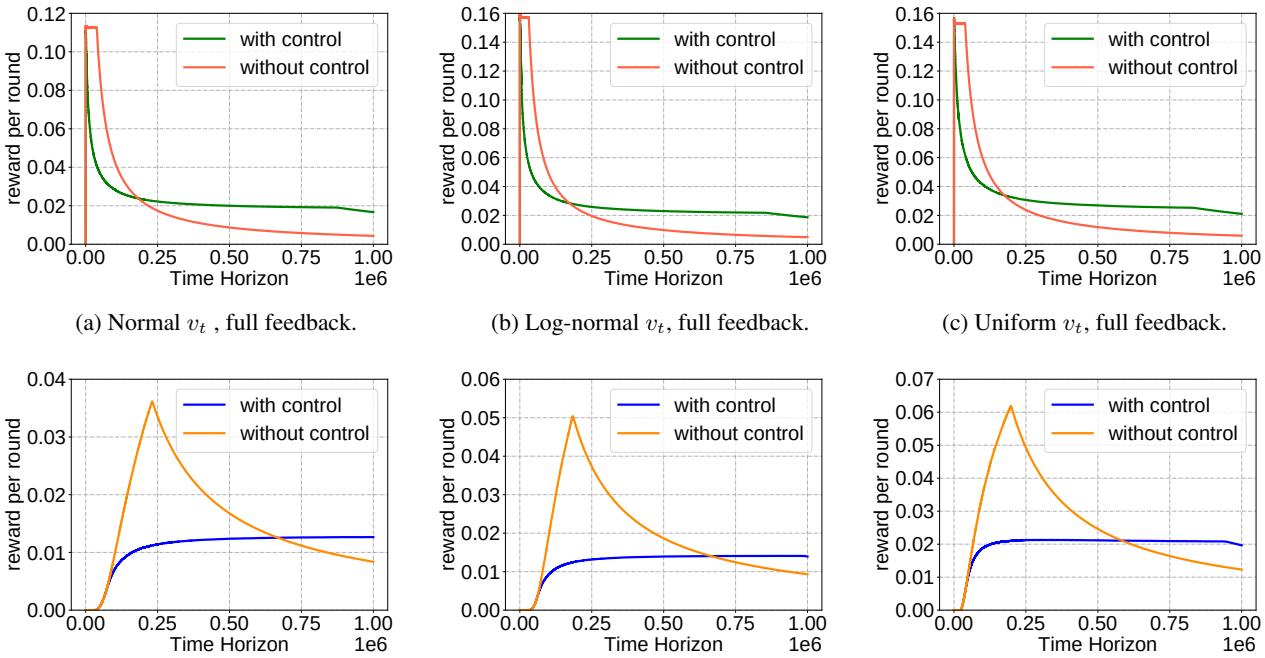
In each experiment, we simulate T rounds auctions under both full and one-sided information structure, and compare our proposed algorithm with ones without budget management, i.e., all the same except for omitting multiplier λ_t and using true value v_t instead of $v_t/(1+\lambda_t)$. The performance is evaluated by observing and plotting $\sum_{s=1}^{t} r_s/t$, the *reward per round* as a function of t. For all algorithms, we uniformly set $M = K = 100$, failure probability $\delta = 0.01$, and adopt fixed step size $\epsilon = 1/\sqrt{T}$. For each of the graph we take the average of 20 independent repetitions of the process.

In the above setup, readers may think $\rho/\bar{v} = 0.01$ is a too tight constraint. However, with an example we show that a seemingly "tight" constraint is necessary for budget management to be of even the least use.

Example 4.1. *In T-round repeated first price auctions with $v_t, d_t \sim \mathcal{U}(0,1)$, when adopting the optimal strategy for Problem (7), the soft budget constraint is not binding if $\rho \geq 1/12$.*

We note that the study on budget management is only needed when budget is relatively tight, such as $\rho < 1/12$ in the above example. Otherwise, a bidder can simply "forget" B, adopt unconstrained strategies, without expecting her budget to run out.

Results and Discussions. The results of three parallel experiments are plotted in Figure 1, with both full and one-sided information feedback considered. Notably, in all in-



(a) Normal v_t , full feedback. (b) Log-normal v_t, full feedback. (c) Uniform v_t, full feedback.

(d) Normal v_t, one-sided feedback. (e) Log-normal v_t, one-sided feedback. (f) Uniform v_t, one-sided feedback.

Figure 1. Performance of bidding algorithms with and without budget control, under full *(Upper)* and one-sided *(Lower)* feedback, evaluated with respect to the reward per round. In three columns, private values are respectively sampled from: *(Left)* normal distribution $v_t \sim \mathcal{N}(0.6, 0.1)$, *(Middle)* logarithmic normal distribution $\log v_t \sim \mathcal{N}(-0.4, 0.1)$, and *(Right)* uniform distribution $v_t \sim \mathcal{U}(0.25, 1)$.

stances, our proposed algorithm outperforms the one with no budget control, with respect to the total reward. The latter algorithm gains remarkable rewards in the beginning rounds, yet tends to deplete its budget in an early phase. The reward per round is then inversely proportional to t, and is eventually exceeded by algorithms with budget control. Meanwhile, for algorithms with budget control, the budget can also be depleted in some instances, but only at the very ending phase, with a delicate turning in the tail of each curve. This coincides our argument that the algorithm has its expected time of budget depletion close to T. We also note that for the bidding algorithm under one-sided information feedback with budget control, $\sum_{s=1}^{t} r_s/t$ holds steady in most rounds, which indicates that the algorithm manages to achieve stable per-round gain as the budget is diminishing. This further demonstrates the effectiveness of the proposed algorithms on budget management.

Further Experiments on Lemma 3.6. We notice that the proposed algorithm performs well in the third experiment where $v_t \sim \mathcal{U}(0.25, 1)$, which does not satisfies the conditions of Assumption 3.5. This indicates that the proposed bidding algorithm with one-sided feedback might perform well on a broader class of private value distributions beyond the requirement of Assumption 3.5. The following experiment provides numerical evidence that Lemma 3.6 is very likely to hold for the value distribution $\mathcal{U}(0.25, 1)$.

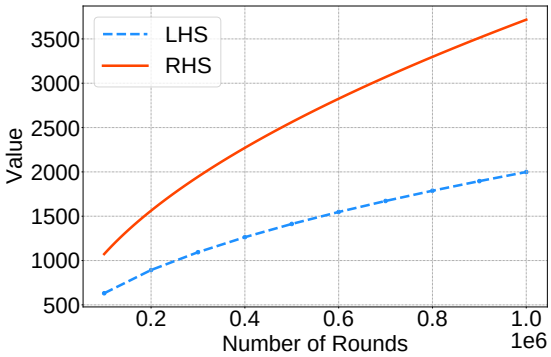


Figure 2. Numerical evidence for Lemma 3.6 by showing $\mathbb{E}\left[\sum_{t=2}^{T}(N_t^{m(t)})^{-1/2}\right] \leq \sqrt{T \ln T}$, in a case where Assumption 3.5 does not hold.

We simulate the bidding algorithm with budget control under one-sided feedback with different time horizons $T = 10^5\tau$ where $\tau = 1, \cdots, 10$, while fixing $K = M = 100$. For each horizon T, we observe the sequence of values and bids to compute $\sum_{t=2}^{T}(N_t^{m(t)})^{-1/2}$. We repeat the process 10 times to estimate its expectation, which is compared with $\sqrt{T \ln T}$. The results are plotted in Figure 2.

In Figure 2, the growth rate of the blue curve is notably smaller than $\sqrt{T \ln T}$, supporting the inequality in Lemma 3.6. We further conjecture that similar properties may hold for a larger class of distributions, only a subset of which is captured by Assumption 3.5. Preciser theoretical characterization of the distribution class is an interesting open problem, which we left as a future direction.

5. Conclusion

In this paper, we study the problem of bidding algorithm design for repeated first-prices auctions with budgets, in both full and one-sided feedback models. On the theoretical side, we prove that Algorithm 1 can achieve an $\widetilde{O}(\sqrt{T})$ regret in the case with full information feedback and that with a technical assumption Algorithm 2 can achieve an $\widetilde{O}(\sqrt{T})$ regret in the case with one-sided information feedback. On the practical side, we show that our algorithms can attain effective budget management as well as good performance. The experiments under distributions that do not satisfy the technical assumption gives evidence that our algorithm can be more widely applicable.

Acknowledgements

This work is supported by Science and Technology Innovation 2030 - "The New Generation of Artificial Intelligence" Major Project (No. 2022ZD0114904). We thank all anonymous reviewers for their helpful feedback.

References

Agrawal, S. and Devanur, N. R. Bandits with global convex constraints and objective. *Operations Research*, 67(5): 1486–1502, 2019.

Ai, R., Wang, C., Li, C., Zhang, J., Huang, W., and Deng, X. No-regret learning in repeated first-price auctions with budget constraints. *arXiv preprint arXiv:2205.14572*, 2022.

Badanidiyuru, A., Kleinberg, R., and Slivkins, A. Bandits with knapsacks. *Journal of the ACM (JACM)*, 65(3):1–55, 2018.

Badanidiyuru, A., Feng, Z., and Guruganesh, G. Learning to bid in contextual first price auctions. *arXiv preprint arXiv:2109.03173*, 2021.

Balseiro, S., Golrezaei, N., Mahdian, M., Mirrokni, V., and Schneider, J. Contextual bandits with cross-learning. *Advances in Neural Information Processing Systems*, 32, 2019.

Balseiro, S. R. and Gur, Y. Learning in repeated auctions with budgets: Regret minimization and equilibrium. *Management Science*, 65(9):3952–3968, 2019.

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.

Balseiro, S. R., Kroer, C., and Kumar, R. Contextual standard auctions with budgets: Revenue equivalence and efficiency guarantees. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pp. 476–476, 2022a.

Balseiro, S. R., Lu, H., and Mirrokni, V. The best of many worlds: Dual mirror descent for online allocation problems. *Operations Research*, 2022b.

Bercu, B. and Touati, A. Exponential inequalities for self-normalized martingales with applications. *The Annals of Applied Probability*, 18(5):1848–1869, 2008.

Bigler, J. Rolling out first price auctions to google ad manager partners. https://www.blog.google/products/admanager/rolling-out-first-price-auctions-google-ad-manager-partners, 2019. Accessed: 08/01/2023.

Capen, E. C., Clapp, R. V., and Campbell, W. M. Competitive bidding in high-risk situations. *Journal of petroleum technology*, 23(06):641–653, 1971.

Castiglioni, M., Celli, A., and Kroer, C. Online learning with knapsacks: the best of both worlds. In *International Conference on Machine Learning*, pp. 2767–2783. PMLR, 2022.

Chen, Z., Wang, C., Wang, Q., Pan, Y., Shi, Z., Tang, C., Cai, Z., Ren, Y., Zhu, Z., and Deng, X. Dynamic budget throttling in repeated second-price auctions. *arXiv preprint arXiv:2207.04690*, 2022.

de la Pena, V. H., Klass, M. J., and Lai, T. L. Self-normalized processes: exponential inequalities, moment bounds and iterated logarithm laws. *Annals of probability*, pp. 1902–1933, 2004.

Despotakis, S., Ravi, R., and Sayedi, A. First-price auctions in online display advertising. *Journal of Marketing Research*, 58(5):888–907, 2021.

Esponda, I. Information feedback in first price auctions. *The RAND Journal of Economics*, 39(2):491–508, 2008.

Feng, Z., Padmanabhan, S., and Wang, D. Online bidding algorithms for return-on-spend constrained advertisers. *arXiv preprint arXiv:2208.13713*, 2022.

Golrezaei, N., Jaillet, P., Liang, J. C. N., and Mirrokni, V. Bidding and pricing in budget and roi constrained markets. *arXiv preprint arXiv:2107.07725*, 2021.

Google Ad Exchange. Bid data sharing. https://support.google.com/authorizedbuyers/answer/2696468, 2022. Accessed: 08/01/2023.

Han, Y., Zhou, Z., Flores, A., Ordentlich, E., and Weissman, T. Learning to bid optimally and efficiently in adversarial first-price auctions. *arXiv preprint arXiv:2007.04568*, 2020a.

Han, Y., Zhou, Z., and Weissman, T. Optimal no-regret learning in repeated first-price auctions. *arXiv preprint arXiv:2003.09795*, 2020b.

Immorlica, N., Sankararaman, K., Schapire, R., and Slivkins, A. Adversarial bandits with knapsacks. *Journal of the ACM*, 69(6):1–47, 2022.

Iyer, K., Johari, R., and Sundararajan, M. Mean field equilibria of dynamic auctions with learning. *Management Science*, 60(12):2949–2970, 2014.

Klemperer, P. *Auctions: theory and practice*. Princeton University Press, 2004.

eMarketer. Worldwide ad spending 2022. https://www.insiderintelligence.com/content/worldwide-ad-spending-2022, 2022. Accessed: 14/12/2022.

Lucking-Reiley, D. Vickrey auctions in practice: From nineteenth-century philately to twenty-first-century e-commerce. *Journal of economic perspectives*, 14(3): 183–192, 2000.

Lucking-Reiley, D., Bryan, D., Prasad, N., and Reeves, D. Pennies from ebay: The determinants of price in online auctions. *The journal of industrial economics*, 55(2): 223–233, 2007.

Massart, P. The tight constant in the dvoretzky-kiefer-wolfowitz inequality. *The annals of Probability*, pp. 1269–1283, 1990.

Zhang, W., Han, Y., Zhou, Z., Flores, A., and Weissman, T. Leveraging the hints: Adaptive bidding in repeated first-price auctions. *arXiv preprint arXiv:2211.06358*, 2022.

A. Missing Proofs in Section 3.1

A.1. Proof of Lemma 3.1

Note that Line 5 in Algorithm 1 essentially estimates $r(v_t, b)$ and $c(b)$ using an empirical distribution $\widetilde{G}_t$:

$$\widetilde{G}_t(b) = \frac{1}{t-1}\sum_{s=1}^{t-1} \mathbf{1}\left\{b \geq d_s\right\}.$$

By Dvoretzky–Kiefer–Wolfowitz (DKW) inequality (Massart, 1990), we have

$$\Pr\left(\sup_b |\widetilde{G}_t(b) - G(b)| \geq \sqrt{\frac{\ln(2T/\delta)}{2(t-1)}}\right) \leq \frac{\delta}{T}.$$

Thus with probability at least $1-\delta$, we have for all $t \geq 2$,

$$\begin{aligned}|\widetilde{r}_t(v_t, b) - r(v_t, b)| \leq& |v_t - b| \cdot |G(b) - \widetilde{G}_t(b)| \leq \bar{v} \cdot \sqrt{\frac{\ln(2T/\delta)}{2(t-1)}},\\ |\widetilde{c}_t(b) - c(b)| \leq& |b| \cdot |G(b) - \widetilde{G}_t(b)| \leq \bar{v} \cdot \sqrt{\frac{\ln(2T/\delta)}{2(t-1)}}.\end{aligned}$$

A.2. Proof of Theorem 3.2

We denote by $\tau \coloneqq \sup\{t \leq T : B_t \geq \bar{v}\}$ the latest period in which the bidder's remaining budget is larger than her maximum private value under Algorithm 1. We consider an alternative framework in which the bidder is allowed to bid even after budget depletion. Note that the performance of π in both the original and alternative frameworks coincide up to time τ. Therefore,

$$R(\pi) = \mathbb{E}_{\boldsymbol{v},\boldsymbol{d}}\left[\sum_{t=2}^{\tau} r_t\right] \geq \mathbb{E}_{\boldsymbol{v},\boldsymbol{d}}\left[\sum_{t=2}^{T} r_t\right] - \bar{v} \cdot \mathbb{E}_{\boldsymbol{v},\boldsymbol{d}}\left[T - \tau\right]. \tag{24}$$

The inequality holds since $r_t \leq v_t \leq \bar{v}$. Here r_t refers to the reward in the alternate framework where the bidder does not break the loop even if $B_{t+1} < \bar{v}$.

We first characterize the optimal strategy for Problem (7). The proof of Lemma A.1 is deferred to Appendix A.3.

Lemma A.1. *There exists an optimal bidding strategy for Problem (7) that maps the value in each round to a random bid, discarding all historical information. Denoting the optimal bidding strategy by α^*, we have*

$$R(\alpha^*) = T \cdot \mathbb{E}^{\alpha^*}_{v\sim F}\left[r(v, \alpha^*(v))\right], \quad \mathbb{E}^{\alpha^*}_{v\sim F}\left[c(\alpha^*(v))\right] \leq \rho.$$

Next we start to lower bound the performance of our strategy. For the first term in the right hand side of (24), we observe that

$$\mathbb{E}^{\pi}_{\boldsymbol{v},\boldsymbol{d}}\left[\sum_{t=2}^{T} r_t\right] = \sum_{t=2}^{T} \mathbb{E}^{\pi}_{\mathcal{H}_t, v_t}\left[r(v_t, b_t)\right] = \sum_{t=2}^{T} \mathbb{E}^{\pi}_{\boldsymbol{v},\boldsymbol{d}}\left[r(v_t, b_t)\right] = \mathbb{E}^{\pi}_{\boldsymbol{v},\boldsymbol{d}}\left[\sum_{t=2}^{T} r(v_t, b_t)\right]. \tag{25}$$

Let α^* be the optimal bidding strategy characterized in Lemma A.1. By the choice of b_t, we have

$$\widetilde{r}_t(v_t, b_t) - \lambda_t \widetilde{c}_t(b_t) \geq \widetilde{r}_t(v_t, \alpha^*(v_t)) - \lambda_t \widetilde{c}_t(\alpha^*(v_t))$$

Then according to Lemma 3.1, with probability at least $1-\delta$, for all $t \geq 2$,

$$r(v_t, b_t) - \lambda_t \widetilde{c}_t(b_t) \geq r(v_t, \alpha^*(v_t)) - \lambda_t c(\alpha^*(v_t)) - (2 + \lambda_t)\bar{v} \cdot \sqrt{\frac{\ln(2T/\delta)}{2(t-1)}}.$$

Reordering terms and summing up from $t = 2$ to T, we have

$$\sum_{t=2}^{T} r(v_t, b_t) \geq \sum_{t=2}^{T} r(v_t, \alpha^*(v_t)) - \sum_{t=2}^{T} \lambda_t c(\alpha^*(v_t)) + \sum_{t=2}^{T} \lambda_t \widetilde{c}_t(b_t) - \sum_{t=2}^{T} (2 + \lambda_t)\bar{v} \cdot \sqrt{\frac{\ln(2T/\delta)}{2(t-1)}}.$$

Notice that the right hand side is upper bounded by $2(T-1)\bar{v}$ since $r(v_t, \alpha^*(v_t)) \leq \bar{v}$ and $\widetilde{c}_t(b_t) \leq 1/(1+\lambda_t)$ by (31).

Taking expectations, we obtain

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{v},\boldsymbol{d}}\left[\sum_{t=2}^{T} r(v_t, b_t)\right] \overset{\text{(a)}}{\geq} & \mathbb{E}_{\boldsymbol{v},\boldsymbol{d}}^{\alpha^*}\left[\sum_{t=2}^{T} r(v_t, \alpha^*(v_t))\right] - \mathbb{E}_{\boldsymbol{v},\boldsymbol{d}}^{\alpha^*}\left[\sum_{t=2}^{T} \lambda_t c(\alpha^*(v_t))\right] \\
& + \mathbb{E}_{\boldsymbol{v},\boldsymbol{d}}\left[\sum_{t=2}^{T} \lambda_t \widetilde{c}_t(b_t)\right] - \left(\frac{\bar{v}^2}{\rho} + \bar{v}\right)\sqrt{2T\ln(2T/\delta)} - \delta \cdot 2(T-1)\bar{v}. \\
\overset{\text{(b)}}{\geq} & R(\alpha^*) - \bar{v} - \mathbb{E}_{\boldsymbol{v},\boldsymbol{d}}\left[\sum_{t=2}^{T} \lambda_t \left(\rho - \widetilde{c}_t(b_t)\right)\right] - \left(\frac{\bar{v}^2}{\rho} + \bar{v}\right)\sqrt{2T\ln(2T/\delta)} - \delta \cdot 2(T-1)\bar{v},
\end{aligned}$$

where (a) follows from $\lambda_t \leq \bar{v}/\rho - 1$ by Lemma A.2 and (b) follows from Lemma A.1.

We now apply a standard analysis of the online gradient descent method to show that the sequence of λ_t is not much worse than a hindsight λ with respect to gain function $h_t(\lambda_t) = \lambda_t\left(\widetilde{c}_t(b) - \rho\right)$. The proof of Lemma A.2 is deferred to Appendix A.4.

Lemma A.2. *For all $t \geq 2$, we have $\lambda_t \in [0, \bar{v}/\rho - 1]$. Moreover, for any $\lambda > 0$, we have*

$$\sum_{t=2}^{T} (\lambda_t - \lambda)\left(\rho - \widetilde{c}_t(b_t)\right) \leq \frac{\lambda^2}{2\epsilon} + \frac{(T-1)\epsilon\bar{v}^2}{2}. \tag{26}$$

By using Lemma A.2 with $\lambda = 0$, we obtain

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{v},\boldsymbol{d}}\left[\sum_{t=2}^{T} r(v_t, b_t)\right] \geq & R(\alpha^*) - \bar{v} - \mathbb{E}_{\boldsymbol{v},\boldsymbol{d}}\left[\sum_{t=2}^{T} \lambda_t \left(\rho - \widetilde{c}_t(b_t)\right)\right] - \left(\frac{\bar{v}^2}{\rho} + \bar{v}\right)\sqrt{2T\ln(2T/\delta)} - \delta \cdot 2(T-1)\bar{v} \\
\geq & R(\alpha^*) - \bar{v} - \frac{(T-1)\epsilon\bar{v}^2}{2} - \left(\frac{\bar{v}^2}{\rho} + \bar{v}\right)\sqrt{2T\ln(2T/\delta)} - \delta \cdot 2(T-1)\bar{v},
\end{aligned} \tag{27}$$

For the second term in the right hand side of (24), we show that the stopping time τ is close to T. The proof of Lemma A.3 is deferred to Appendix A.5

Lemma A.3. *For Algorithm 1, with probability at least $1 - 2\delta$, we have*

$$T - \tau \leq \frac{\bar{v}}{\rho} \cdot \left(\frac{1}{\epsilon\rho} + \sqrt{2T\ln(2T/\delta)} + \sqrt{2T\ln(1/\delta)}\right).$$

By Lemma A.3, we have

$$\mathbb{E}_{\boldsymbol{v},\boldsymbol{d}}[T - \tau] \leq (1 - 2\delta) \cdot \frac{\bar{v}}{\rho} \cdot \left(\frac{1}{\epsilon\rho} + \sqrt{2T\ln(2T/\delta)} + \sqrt{2T\ln(1/\delta)}\right) + 2\delta \cdot T. \tag{28}$$

Plugging (27) and (28) into (24), we obtain

$$\begin{aligned}
R(\pi) \geq & R(\alpha^*) - \bar{v} - \frac{(T-1)\epsilon\bar{v}^2}{2} - \left(\frac{\bar{v}^2}{\rho} + \bar{v}\right)\sqrt{2T\ln(2T/\delta)} - \delta \cdot 2(T-1)\bar{v} \\
& - (1 - 2\delta) \cdot \frac{\bar{v}^2}{\rho} \cdot \left(\frac{1}{\epsilon\rho} + \sqrt{2T\ln(2T/\delta)} + \sqrt{2T\ln(1/\delta)}\right) - 2\delta \cdot T\bar{v}.
\end{aligned}$$

By setting the step size to $\epsilon \sim T^{-1/2}$ and the failure probability to $\delta \sim T^{-1}$, Algorithm 1 can obtain a regret of order $O(\sqrt{T\log T})$.

A.3. Proof of Lemma A.1

Let $\alpha : [0, \bar{v}] \mapsto \Delta[0, \bar{v}]$ be a bidding strategy that maps v_t to a distribution over $[0, \bar{v}]$. Consider the following optimization problem:

$$\max_{\alpha} \mathbb{E}^{\alpha}_{\boldsymbol{v},\boldsymbol{d}} \left[\sum_{t=1}^{T} \mathbf{1}\left\{\alpha(v_t) \geq d_t\right\} (v_t - \alpha(v_t)) \right]$$
$$\text{s.t. } \mathbb{E}^{\alpha}_{\boldsymbol{v},\boldsymbol{d}} \left[\sum_{t=1}^{T} \mathbf{1}\left\{\alpha(v_t) \geq d_t\right\} \alpha(v_t) \right] \leq \rho T.$$

Because the sequences of v_t and d_t are independent samples, the problem can be simplified as

$$\begin{aligned} &\max_{\alpha} T \cdot \mathbb{E}^{\alpha}_{v\sim F} \left[(v - \alpha(v))\, G(\alpha(v)) \right] \\ &\text{s.t. } \mathbb{E}^{\alpha}_{v\sim F} \left[\alpha(v) G(\alpha(v)) \right] \leq \rho. \end{aligned} \tag{29}$$

Let $\mathcal{A}_1$ be the set of all feasible solutions to Problem (29). Note that this is a convex optimization problem where Slater's condition holds (always bidding 0 is an interior point). As a result, strong duality holds:

$$\max_{\alpha\in\mathcal{A}_1} R(\alpha) = \min_{\lambda\geq 0} T \cdot \left(\mathbb{E}_v \left[\max_{\alpha} \left(v - (1+\lambda)\alpha(v)\right) G\left(\alpha(v)\right) \right] + \lambda\rho \right).$$

The optimal bidding strategy α^* maps v to a distribution over the bids that maximize $(v - (1+\lambda^*)b)\, G(b)$, where λ^* satisfies the complementary conditions

$$\lambda^* \geq 0 \perp \mathbb{E}^{\alpha^*}_{v\sim F} \left[\alpha^*(v) G(\alpha^*(v)) \right] \leq \rho.$$

On the one hand, we have $\mathcal{A}_1 \subseteq \Pi_1$ so $\max_{\alpha\in\mathcal{A}_1} R(\alpha) \leq \max_{\pi\in\Pi_1} R(\pi)$. On the other hand, the performance of strategy α^* achieves the right hand side of (8), an upper bound of $\max_{\pi\in\Pi_1} R(\pi)$. Therefore, α^* is also an optimal strategy for Problem (7).

A.4. Proof of Lemma A.2

By the choice of bid in Line 6, we have

$$\widetilde{r}_t(v_t, b_t) - \lambda_t \widetilde{c}_t(b_t) \geq \widetilde{r}_t(v_t, 0) - \lambda_t \widetilde{c}_t(0) \geq 0, \tag{30}$$

and then,

$$\begin{aligned} (1+\lambda_t)\widetilde{c}_t(b_t) \leq \widetilde{r}_t(v_t, b_t) + \widetilde{c}_t(b_t) \leq & \frac{1}{t-1} \sum_{s=1}^{t-1} \mathbf{1}\left\{b \geq d_s\right\} v_t \leq v_t \leq \bar{v} \\ \implies \widetilde{c}_t(b_t) \leq & \frac{\bar{v}}{1+\lambda_t}. \end{aligned} \tag{31}$$

Meanwhile, inequality (30) implies $b_t \leq v_t/(1+\lambda_t)$, otherwise $\widetilde{r}_t(v_t, b_t) - \lambda_t \widetilde{c}_t(b_t) \leq 0 \leq \widetilde{r}_t(v_t, 0) - \lambda_t \widetilde{c}_t(0)$, which means that the algorithm should have chosen a smaller bid instead.

According to the update rule of λ_t, when $\lambda_t \leq \bar{v}/\rho - 1$, we have

$$\lambda_t - \epsilon\left(\rho - \widetilde{c}_t(b_t)\right) \overset{\text{(a)}}{\leq} \lambda_t + \frac{\epsilon\bar{v}}{1+\lambda_t} - \epsilon\rho \overset{\text{(b)}}{\leq} \max\{\epsilon\bar{v} - \epsilon\rho, \frac{\bar{v}}{\rho} - 1\} \overset{\text{(c)}}{=} \frac{\bar{v}}{\rho} - 1,$$

where (a) follows from inequality (31), (b) follows from that $\phi(x) = x + \epsilon\bar{v}/(1+x)$ is convex over $\mathbb{R}_+$, and (c) holds since $\epsilon = 1/\sqrt{T} < 1/\rho$. Because we take $\lambda_2 = 0$ in initialization, by induction, we have $\lambda_t \in [0, \bar{v}/\rho - 1]$ for all $t \geq 2$.

Again by the update rule of λ_t in Line 7, we have for any $\lambda \geq 0$,

$$\begin{aligned} \|\lambda_{t+1} - \lambda\|_2^2 \overset{\text{(a)}}{\leq} & \|\lambda_t - \epsilon(\rho - \widetilde{c}_t(b_t)) - \lambda\|_2^2, \\ = & \|\lambda_t - \lambda\|^2 - 2\epsilon\left(\lambda_t - \lambda\right)\left(\rho - \widetilde{c}_t(b_t)\right) + \epsilon^2 \|\rho - \widetilde{c}_t(b_t)\|_2^2 \\ \overset{\text{(b)}}{\leq} & \|\lambda_t - \lambda\|^2 - 2\epsilon\left(\lambda_t - \lambda\right)\left(\rho - \widetilde{c}_t(b_t)\right) + \epsilon^2 \bar{v}^2, \end{aligned}$$

where (a) follows from a standard contraction property of projection operator and (b) holds by $\rho \leq \bar{v}$ and (31).

Reordering terms and summing up from $t = 2$ to T, we have

$$\sum_{t=2}^{T} (\lambda_t - \lambda)(\rho - \widetilde{c}_t(b_t)) \leq \frac{\|\lambda_2 - \lambda\|_2^2 - \|\lambda_{T+1} - \lambda\|_2^2}{2\epsilon} + \frac{(T-1)\epsilon\bar{v}^2}{2}.$$

which leads to (26) with $\lambda_2 = 0$.

A.5. Proof of Lemma A.3

Reordering $\lambda_{t+1} \geq \lambda_t - \epsilon(\rho - \widetilde{c}_t(b_t))$ and summing over $t = 2, \ldots, \tau$, we have

$$\sum_{t=2}^{\tau} (\widetilde{c}_t(b_t) - \rho) \leq \frac{\lambda_{\tau+1}}{\epsilon} \leq \frac{\bar{v}/\rho - 1}{\epsilon}. \tag{32}$$

For the left hand side of (32), we use inequality (10). With probability at least $1 - \delta$,

$$\sum_{t=2}^{\tau} (\widetilde{c}_t(b_t) - \rho) \geq \sum_{t=2}^{\tau} c(b_t) - \tau\rho - \sum_{t=2}^{\tau} \bar{v} \cdot \sqrt{\frac{\ln(2T/\delta)}{2(t-1)}} \tag{33}$$

Let $X_t := \sum_{s=2}^{t}(c_s - c(b_s))$. Because $\mathbb{E}_{d_t}[c_t - c(b_t)|\mathcal{H}_t] = 0$, we know that$\{X_1, X_2, \ldots, X_\tau\}$ is a martingale. Applying Azuma-Hoeffding inequality, we have the following inequality holds with failure probability at most δ,

$$\sum_{t=2}^{\tau} (c_t - c(b_t)) \leq \bar{v} \cdot \sqrt{2T\ln(1/\delta)}. \tag{34}$$

According to the definition of τ, when $\tau < T$,

$$B_{\tau+1} < \bar{v} \Longrightarrow \sum_{t=2}^{\tau} c_t > \rho T - \bar{v}. \tag{35}$$

Combining (32), (33), (34) and (35), we obtain with probability at least $1 - 2\delta$,

$$\begin{aligned}\rho(T - \tau) \leq & \frac{\bar{v}/\rho - 1}{\epsilon} + \sum_{t=2}^{\tau} \bar{v} \cdot \sqrt{\frac{\ln(2T/\delta)}{2(t-1)}} + \bar{v} \cdot \sqrt{2T\ln(1/\delta)} + \bar{v} \\ \leq & \bar{v} \cdot \left(\frac{1}{\epsilon\rho} + \sqrt{2T\ln(2T/\delta)} + \sqrt{2T\ln(1/\delta)}\right),\end{aligned} \tag{36}$$

Note that when $\tau = T$, the inequality holds trivially.

B. Missing Proofs in Section 3.2

B.1. Proof of Lemma 3.4

By definition, we have

$$\begin{aligned}&\widetilde{r}_t(v^m, b^k) - r(v^m, b^k) \\ &= \left(v^m - b^k\right) \cdot \frac{\sum_{s=1}^{t-1} \mathbf{1}\left\{b_s \leq b^k\right\}\left(\mathbf{1}\left\{b^k \geq d_s\right\} - G(b^k)\right)}{\sum_{s=1}^{t-1} \mathbf{1}\left\{b_s \leq b^k\right\}}.\end{aligned}$$

We denote the above numerator by X_t. As $\mathbb{E}[X_{t+1} - X_t|\mathcal{H}_t] = 0$, the sequence of X_t is a martingale adapted to the filtration $\{\mathcal{H}_1, \mathcal{H}_2, \ldots\}$. Next we make use the following two lemmas.

Lemma B.1 (Bercu & Touati (2008)). *Let $\{X_1, X_2, \dots\}$ be a locally square integrable martingale. Denote*

$$V_t(\mu) = \exp\left(\mu X_t - \frac{\mu^2}{2}\left(\langle X\rangle_t + [X]_t\right)\right),$$

where the predictable quadratic variation $\langle X\rangle_t$ and the total quadratic variation $[X]_t$ are respectively defined by

$$\langle X\rangle_t = \sum_{s=1}^{t-1} \mathbb{E}\left[(X_s - X_{s-1})^2 \,|\mathcal{H}_{s-1}\right], \quad [X]_t = \sum_{s=1}^{t-1} (X_s - X_{s-1})^2.$$

Then, for any μ, the sequence of $V_t(\mu)$ is a positive super-martingale with $\mathbb{E}\left[V_t(\mu)\right] \le 1$.

Lemma B.2 (de la Pena et al. (2004)). *Let A and $B \ge 0$ be two random variables satisfying for any μ,*

$$\mathbb{E}\left[\mu A - \frac{\mu^2}{2} B^2\right] \le 1.$$

The for any $x \ge \sqrt{2}$, $y > 0$, we have

$$\Pr\left(|A| \Big/ \sqrt{(B^2+y)\left(1 + \frac{1}{2}\ln\left(\frac{B^2}{y}+1\right)\right)} \ge x\right) \le \exp\left(-\frac{x^2}{2}\right)$$

By Lemma B.1, $A = X_t$ and $B = \sqrt{\langle X\rangle_t + [X]_t}$ satisfies the conditions of Lemma B.2. Taking $x = \sqrt{2\ln(KT/\delta)}$ and $y = 1$, we obtain

$$\Pr\left(\frac{|A|}{\sqrt{(B^2+1)\left(2+\ln\left(B^2+1\right)\right)}} \ge \sqrt{\ln(KT/\delta)}\right) \le \frac{\delta}{KT}.$$

Next, it holds that

$$\begin{aligned}\left(B^2+1\right)\left(2+\ln\left(B^2+1\right)\right) &\overset{\text{(a)}}{\le} \left(2n_t^k+1\right)\left(2+\ln\left(2n_t^k+1\right)\right)\\ &\overset{\text{(b)}}{\le} n_t^k\left(6+3\ln(2t-1)\right)\\ &\overset{\text{(c)}}{\le} 4n_t^k \ln T.\end{aligned}$$

where (a) holds because $B^2 = \langle X\rangle_t + [X]_t \le 2n_t^k$, (b) holds since $n_t^k \in [1, t-1]$ (note that $b_1 = 0$), and (c) holds when T is sufficiently large.

Thus, we have with probability at least $1-\delta$, $\forall t \ge 2, m \in [M], k \in [K]$,

$$|\widetilde{r}_t(v^m, b^k) - r(v^m, b^k)| \le \bar{v} \cdot \sqrt{\frac{4\ln T \ln(KT/\delta)}{n_t^k}}.$$

The same analysis goes for $\widetilde{c}_t(b^k)$.

B.2. Proof of Lemma 3.6

We first prove that for any $t \ge 2$,

$$N_t^{m(t)} \ge 1 + \sum_{s=2}^{t-1} \mathbf{1}\left\{\frac{v_s}{1+\lambda_s} \le \frac{v_t}{1+\lambda_t}\right\}. \tag{37}$$

Note that the bidder always bid 0 in the first round so $N_t^{m(t)} \ge 1$. For every past round $2 \le s < t$ with $v_s/(1+\lambda_s) \le v_t/(1+\lambda_t)$, as $v^{m(s)} \le v^{m(t)}$, Line 7 in Algorithm 2 guarantees that $\inf \mathcal{B}_s^{m(s)} \le \inf \mathcal{B}_s^{m(t)}$. Also notice that $\mathcal{B}_{t-1}^{m(t)} \subseteq \mathcal{B}_s^{m(t)}$ by the bid elimination rule. Therefore, we have for all $b^k \in \mathcal{B}_{t-1}^{m(t)}$,

$$b_s = \inf \mathcal{B}_s^{m(s)} \le \inf \mathcal{B}_{t-1}^{m(t)} \le b^k. \tag{38}$$

By definition,

$$N_t^{m(t)} = \min_{b^k \in \mathcal{B}_{t-1}^{m(t)}} n_t^k = \min_{b^k \in \mathcal{B}_{t-1}^{m(t)}} \sum_{s=1}^{t-1} \mathbf{1}\left\{b_s \le b^k\right\}.$$

The inequality (38) implies that every past value v_s with $v_s/(1+\lambda_s) \le v_t/(1+\lambda_t)$ has contributed to the value of $N_t^{m(t)}$ by 1, which leads to the result of (37).

Next, due to $\lambda_t \le \bar{v}/\rho - 1$ by Lemma A.2, we further have $N_t^{m(t)} \ge 1 + \sum_{s=2}^{t-1} \mathbf{1}\left\{v_s \le (\rho/\bar{v})v_t\right\}$. Then,

$$\mathbb{E}_{\boldsymbol{v},\boldsymbol{d}}\left[\sum_{t=2}^{T}\sqrt{1/N_t^{m(t)}}\right] \le \sum_{t=2}^{T}\mathbb{E}_{\boldsymbol{v},\boldsymbol{d}}\left[\sqrt{\frac{1}{1+\sum_{s=2}^{t-1}\mathbf{1}\left\{v_s \le (\rho/\bar{v})v_t\right\}}}\right] \le \sum_{t=2}^{T}\sqrt{\mathbb{E}_{\boldsymbol{v},\boldsymbol{d}}\left[\frac{1}{1+\sum_{s=2}^{t-1}\mathbf{1}\left\{v_s \le (\rho/\bar{v})v_t\right\}}\right]},$$

where the last equality follows from $\mathbb{E}[y^2] \ge (\mathbb{E}[y])^2$ for any random variable y.

Conditioned on v_t, the sum $\sum_{s=2}^{t-1} \mathbf{1}\left\{v_s \le (\rho/\bar{v})v_t\right\}$ follows a binomial distribution $\mathrm{Binomial}(t-2, F((\rho/\bar{v})v_t))$. Thus,

$$\begin{aligned}
\mathbb{E}_{v_1,\ldots,v_{t-1}}\left[\frac{1}{1+\sum_{s=2}^{t-1}\mathbf{1}\left\{v_s \le (\rho/\bar{v})v_t\right\}}\right] =& \sum_{s=0}^{t-2}\frac{1}{1+s}\binom{t-2}{s}F^s((\rho/\bar{v})v_t)\left(1-F((\rho/\bar{v})v_t)\right)^{t-2-s} \\
=& \sum_{s=0}^{t-2}\frac{1}{t-1}\binom{t-1}{s+1}F^s((\rho/\bar{v})v_t)\left(1-F((\rho/\bar{v})v_t)\right)^{t-2-s} \\
=& \frac{1}{t-1}\sum_{s=1}^{t-1}\binom{t-1}{s}F^{s-1}((\rho/\bar{v})v_t)\left(1-F((\rho/\bar{v})v_t)\right)^{t-1-s} \\
=& \frac{1}{(t-1)F((\rho/\bar{v})v_t)}\left(1-\left(1-F((\rho/\bar{v})v_t)\right)^{t-1}\right) \\
\le& \frac{1}{(t-1)F((\rho/\bar{v})v_t)}.
\end{aligned}$$

Note that the conditional expectation is also upper bounded by 1 since $\sum_{s=2}^{t-1} \mathbf{1}\left\{v_s \le (\rho/\bar{v})v_t\right\} \ge 0$. Consequently, we have

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{v},\boldsymbol{d}}\left[\frac{1}{1+\sum_{s=2}^{t-1}\mathbf{1}\left\{v_s \le (\rho/\bar{v})v_t\right\}}\right] \le& \mathbb{E}_{v_t}\left[\min\left\{\frac{1}{(t-1)F((\rho/\bar{v})v_t)}, 1\right\}\right] \\
=& \int_{(t-1)F((\rho/\bar{v})v_t)\ge 1}\frac{1}{(t-1)F((\rho/\bar{v})v_t)}\,dF(v_t) + \int_{(t-1)F((\rho/\bar{v})v_t)\le 1} dF(v_t) \\
=& \left.\frac{F(v_t)}{(t-1)F((\rho/\bar{v})v_t)}\right|_{(\bar{v}/\rho)F^{-1}(1/(t-1))}^{\bar{v}} \\
&- \int_{(\bar{v}/\rho)F^{-1}(1/(t-1))}^{\bar{v}} F(v_t)\,d\left(\frac{1}{(t-1)F((\rho/\bar{v})v_t)}\right) + F((\bar{v}/\rho)F^{-1}(1/(t-1))) \\
=& \frac{1}{(t-1)F(\rho)} + \int_{(\bar{v}/\rho)F^{-1}(1/(t-1))}^{\bar{v}} \frac{(\rho/\bar{v})F(v_t)f((\rho/\bar{v})v_t)}{(t-1)F^2((\rho/\bar{v})v_t)}\,d(v_t) \\
\le& \frac{1}{(t-1)F(\rho)} + \frac{\overline{f}\bar{v}}{(t-1)\underline{f}\rho}\cdot \left.\ln F((\rho/\bar{v})v_t)\right|_{(\bar{v}/\rho)F^{-1}(1/(t-1))}^{\bar{v}} \\
=& \underbrace{\left(\frac{1}{F(\rho)} + \frac{\overline{f}\bar{v}\ln F(\rho)}{\underline{f}\rho}\right)}_{C_1}\cdot\frac{1}{t-1} + \underbrace{\frac{\overline{f}\bar{v}}{\underline{f}\rho}}_{C_2}\cdot\frac{\ln(t-1)}{t-1}
\end{aligned}$$

Summing up the square-roots over $t = 2, \dots, T$, we have

$$\begin{aligned}\mathbb{E}_{\boldsymbol{v},\boldsymbol{d}}\left[\sum_{t=2}^{T}\sqrt{1/N_t^{m(t)}}\right] \leq& \sum_{t=2}^{T}\sqrt{C_1\cdot\frac{1}{t-1}+C_2\frac{\ln(t-1)}{t-1}}\\ \leq& \sqrt{C_1}\sum_{t=1}^{T-1}\sqrt{\frac{1}{t}}+\sqrt{C_2}\sum_{t=1}^{T-1}\sqrt{\frac{\ln t}{t}}\\ \leq& \left(\sqrt{C_1}+\sqrt{C_2\ln T}\right)\sum_{t=1}^{T-1}t^{-1/2}\\ \leq& 2\left(\sqrt{C_1}+\sqrt{C_2\ln T}\right)\sqrt{T},\end{aligned}$$

where the second inequality follows from $\sqrt{x+y} \leq \sqrt{x}+\sqrt{y}$.

B.3. Proof of Theorem 3.7

Similar to the proof of Theorem 3.2, we denote by $\tau := \sup\{t \leq T : B_t \geq \bar{v}\}$ the last round before budget depletion under Algorithm 2 and consider an alternative framework in which the bidder is allowed to bid even after budget depletion. Then, we have

$$R(\pi) = \mathbb{E}_{\boldsymbol{v},\boldsymbol{d}}\left[\sum_{t=2}^{\tau} r_t\right] \geq \mathbb{E}_{\boldsymbol{v},\boldsymbol{d}}\left[\sum_{t=2}^{T} r_t\right] - \bar{v}\cdot\mathbb{E}_{\boldsymbol{v},\boldsymbol{d}}\left[T-\tau\right] = \mathbb{E}^{\pi}_{\boldsymbol{v},\boldsymbol{d}}\left[\sum_{t=2}^{T} r(v_t,b_t)\right] - \bar{v}\cdot\mathbb{E}_{\boldsymbol{v},\boldsymbol{d}}\left[T-\tau\right], \tag{39}$$

where the last equality follows equation (25).

First, we make use of the following lemma, the proof of which is deferred to Appendix B.4

Lemma B.3. *Let $\widetilde{b}(v) = \arg\max_{b\in\mathcal{B}}(v-b)G(b)$ (taking the smallest b if there are ties). Then for Algorithm 2, with probability at least $1-\delta$, $\forall t \geq 2, m \in [M]$, $\widetilde{b}(v^m) \in B_t^m$.*

Then with probability at least $1-\delta$, for all t,

$$\begin{aligned}r(v^{m(t)},b_t) \geq& \widetilde{r}_t(v^{m(t)},b_t) - w_t^{m(t)}\\ \geq& \widetilde{r}_t(v^{m(t)},\widetilde{b}(v^{m(t)})) - 3w_t^{m(t)}\\ \geq& r(v^{m(t)},\widetilde{b}(v^{m(t)})) - 4w_t^{m(t)},\end{aligned} \tag{40}$$

where the first and third inequalities follows from Lemma 3.4, and the second inequality holds by Lemma B.3.

Next, for the left hand side of (40), we have

$$\begin{aligned}r(v^{m(t)},b_t) =& (v^{m(t)}-b_t)G(b_t)\\ =& \left(\frac{v_t}{1+\lambda_t}-b_t\right)G(b_t) - \left(\frac{v_t}{1+\lambda_t}-v^{m(t)}\right)G(b_t)\\ =& \frac{1}{1+\lambda_t}\left(r(v_t,b_t)-\lambda_t c(b_t)\right) - \left(\frac{v_t}{1+\lambda_t}-v^{m(t)}\right)G(b_t).\end{aligned} \tag{41}$$

Let $b^*(v) = \arg\max_b(v-b)G(b)$ (taking the smallest b if there are ties) and $b'(v) = \min\{b \in \mathcal{B} : b \geq b^*(v)\}$. For the

right hand side of (40), we have

$$\begin{aligned}
r(v^{m(t)}, \widetilde{b}(v^{m(t)})) \overset{\text{(a)}}{\geq} & r(v^{m(t)}, b'(v^{m(t)})) \\
= & (v^{m(t)} - b'(v^{m(t)}))G(b'(v^{m(t)})) \\
\overset{\text{(b)}}{\geq} & (v^{m(t)} - b'(v^{m(t)}))G(b^*(v^{m(t)})) \\
\overset{\text{(c)}}{\geq} & (v^{m(t)} - b^*(v^{m(t)}))G(b^*(v^{m(t)})) - \frac{\bar{v}}{K} \\
= & r(v^{m(t)}, b^*(v^{m(t)})) - \frac{\bar{v}}{K},
\end{aligned} \tag{42}$$

where (a) and (b) hold by the definition of $\widetilde{b}(v)$, and (c) holds since $b'(v) \leq b^*(v) + \bar{v}/K$.

Let α^* be the optimal bidding strategy characterized in Lemma A.1. By the definition of $b^*(v)$, We have

$$\begin{aligned}
r(v^{m(t)}, b^*(v^{m(t)})) \geq & r(v^{m(t)}, \alpha^*(v_t)) \\
= & (v^{m(t)} - \alpha^*(v_t))G(\alpha^*(v_t)) \\
= & \left(\frac{v_t}{1+\lambda_t} - \alpha^*(v_t)\right) G(\alpha^*(v_t)) - \left(\frac{v_t}{1+\lambda_t} - v^{m(t)}\right) G(\alpha^*(v_t)) \\
= & \frac{1}{1+\lambda_t}\left(r(v_t, \alpha^*(v_t)) - \lambda_t c(\alpha^*(v_t))\right) - \left(\frac{v_t}{1+\lambda_t} - v^{m(t)}\right) G(\alpha^*(v_t)).
\end{aligned} \tag{43}$$

Putting (40), (41), (42) and (43) together, we obtain

$$\begin{aligned}
r(v_t, b_t) - \lambda_t c(b_t) \geq & r(v_t, \alpha^*(v_t)) - \lambda_t c(\alpha^*(v_t)) - (1+\lambda_t)\left(\frac{\bar{v}}{K} + 4w_t^m\right) \\
& + (1+\lambda_t)\left(\frac{v_t}{1+\lambda_t} - v^{m(t)}\right)(G(b_t) - G(\alpha^*(v_t))) \\
\geq & r(v_t, \alpha^*(v_t)) - \lambda_t c(\alpha^*(v_t)) - (1+\lambda_t)\left(\frac{\bar{v}}{K} + \frac{\bar{v}}{M} + 4w_t^{m(t)}\right),
\end{aligned}$$

where the second inequality holds since $v_t/(1+\lambda_t) - v^{m(t)} \leq \bar{v}/M$. Reordering terms and summing up from $t = 2$ to T, we have with probability at least $1 - \delta$,

$$\begin{aligned}
\sum_{t=2}^{T} r(v_t, b_t) \geq & \sum_{t=2}^{T} r(v_t, \alpha^*(v_t)) - \sum_{t=2}^{T} \lambda_t c(\alpha^*(v_t)) - \sum_{t=2}^{T}(1+\lambda_t)\left(\frac{\bar{v}}{K} + \frac{\bar{v}}{M} + 4w_t^{m(t)}\right) + \sum_{t=2}^{T} \lambda_t c(b_t) \\
\geq & \sum_{t=2}^{T} r(v_t, \alpha^*(v_t)) - \sum_{t=2}^{T} \lambda_t c(\alpha^*(v_t)) - \sum_{t=2}^{T}(1+\lambda_t)\left(\frac{\bar{v}}{K} + \frac{\bar{v}}{M} + 4w_t^{m(t)}\right) + \sum_{t=2}^{T} \lambda_t \widetilde{c}_t(b_t) - \sum_{t=2}^{T} \lambda_t w_t^{m(t)},
\end{aligned}$$

where we further apply Lemma 3.4 for the second inequality.

Taking expectations, we obtain

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{v},\boldsymbol{d}}\left[\sum_{t=2}^{T} r(v_t, b_t)\right] \geq & \mathbb{E}^{\alpha^*}_{\boldsymbol{v},\boldsymbol{d}}\left[\sum_{t=2}^{T} r(v_t, \alpha^*(v_t))\right] - \mathbb{E}^{\alpha^*}_{\boldsymbol{v},\boldsymbol{d}}\left[\sum_{t=2}^{T} \lambda_t c(\alpha^*(v_t))\right] - \frac{\bar{v}^2}{\rho}(T-1)\left(\frac{1}{K}+\frac{1}{M}\right) \\
& + \mathbb{E}_{\boldsymbol{v},\boldsymbol{d}}\left[\sum_{t=2}^{T} \lambda_t \widetilde{c}_t(b_t)\right] - \left(\frac{5\bar{v}}{\rho}-1\right)\mathbb{E}_{\boldsymbol{v},\boldsymbol{d}}\left[\sum_{t=2}^{T} w_t^{m(t)}\right] - \delta\cdot 2(T-1)\bar{v} \\
\overset{\text{(a)}}{\geq} & R(\alpha^*) - \bar{v} - \mathbb{E}_{\boldsymbol{v},\boldsymbol{d}}\left[\sum_{t=2}^{T} \lambda_t\left(\rho - \widetilde{c}_t(b_t)\right)\right] - \frac{\bar{v}^2}{\rho}(T-1)\left(\frac{1}{K}+\frac{1}{M}\right) \\
& - \left(\frac{5\bar{v}}{\rho}-1\right)\mathbb{E}_{\boldsymbol{v},\boldsymbol{d}}\left[\sum_{t=2}^{T} w_t^{m(t)}\right] - \delta\cdot 2(T-1)\bar{v} \\
\overset{\text{(a)}}{\geq} & R(\alpha^*) - \bar{v} - \frac{(T-1)\epsilon\bar{v}^2}{2} - \frac{\bar{v}^2}{\rho}(T-1)\left(\frac{1}{K}+\frac{1}{M}\right) \\
& - \left(\frac{5\bar{v}}{\rho}-1\right)\mathbb{E}_{\boldsymbol{v},\boldsymbol{d}}\left[\sum_{t=2}^{T} w_t^{m(t)}\right] - \delta\cdot 2(T-1)\bar{v},
\end{aligned} \tag{44}$$

where (a) follows from Lemma A.1 and $\lambda_t \leq \bar{v}/\rho - 1$ in Lemma A.2, (b) holds by using Lemma A.2 with $\lambda = 0$. Remark that Lemma A.2 still holds for Algorithm 2 under one-sided information feedback.

Next, we provide a result for one-sided information feedback analogous to Lemma A.3.

Lemma B.4. *For Algorithm 2, with probability at least $1-2\delta$, we have*

$$T - \tau \leq \frac{\bar{v}}{\rho}\cdot\left(\frac{1}{\epsilon\rho} + \sqrt{2T\ln(1/\delta)}\right) + \frac{1}{\rho}\mathbb{E}_{\boldsymbol{v},\boldsymbol{d}}\left[\sum_{t=2}^{T} w_t^{m(t)}\right].$$

The proof is exactly the same to that of Lemma A.3, except that we use Lemma 3.4 rather than Lemma 3.1 to bound the estimation error. We omit the proof.

By Lemma B.4, we have

$$\mathbb{E}_{\boldsymbol{v},\boldsymbol{d}}\left[T-\tau\right] \leq (1-2\delta)\cdot\frac{\bar{v}}{\rho}\cdot\left(\frac{1}{\epsilon\rho} + \sqrt{2T\ln(1/\delta)}\right) + (1-2\delta)\cdot\frac{1}{\rho}\mathbb{E}_{\boldsymbol{v},\boldsymbol{d}}\left[\sum_{t=2}^{T} w_t^{m(t)}\right] + 2\delta\cdot T \tag{45}$$

Plugging (44) and (45) into (39), we obtain

$$\begin{aligned}
R(\pi) \geq & R(\alpha^*) - \bar{v} - \frac{(T-1)\epsilon\bar{v}^2}{2} - \frac{\bar{v}^2}{\rho}(T-1)\left(\frac{1}{K}+\frac{1}{M}\right) \\
& - \delta\cdot 2(T-1)\bar{v} - (1-2\delta)\cdot\frac{\bar{v}^2}{\rho}\cdot\left(\frac{1}{\epsilon\rho} + \sqrt{2T\ln(1/\delta)}\right) - 2\delta\cdot T\bar{v} \\
& - \left(\frac{5\bar{v}^2}{\rho} - \bar{v} + (1-2\delta)\frac{\bar{v}^2}{\rho}\right)\sqrt{4\ln T\log(KT/\delta)}\,\mathbb{E}_{\boldsymbol{v},\boldsymbol{d}}\left[\sum_{t=2}^{T}\sqrt{\frac{1}{N_t^{m(t)}}}\right].
\end{aligned}$$

By setting the step size to $\epsilon \sim T^{-1/2}$ and the failure probability to $\delta \sim T^{-1}$, together with Lemma 3.6, we can obtain the desired regret bound.

B.4. Proof of Lemma B.3

We first prove that $\widetilde{b}(v)$ is non-decreasing in v. Let v^s and v^m be two values with $v^s \leq v^m$. Then for any $b \in \mathcal{B}$ with $b \geq \widetilde{b}(v^m)$, we have

$$\begin{aligned} r(v^s, b) =& (v^s - b)G(b) = (v^m - b)G(b) - (v^m - v^s)G(b) \\ \leq& (v^m - \widetilde{b}(v^m))G(b^*(v^m)) - (v^m - v^s)G(\widetilde{b}(v^m)) \\ =& (v^s - \widetilde{b}(v^m))G(\widetilde{b}(v^m)) = r(v^s, \widetilde{b}(v^m)). \end{aligned}$$

Thus, it holds that $\widetilde{b}(v^s) \leq \widetilde{b}(v^m)$ by the definition of $\widetilde{b}(v)$.

Next, we prove the result by induction. Suppose that for all $s < m$, $\widetilde{b}(v^s) \in B_t^s$ in round t. By the non-decreasing property of $\widetilde{b}(v)$, we have

$$\widetilde{b}(v^m) \geq \widetilde{b}(v^s) \geq \inf B_t^s.$$

Thus, $\widetilde{b}(v^m)$ is not eliminated by Line 7. As for Line 9, we have for any $k \in \mathcal{B}_{t-1}^m$,

$$\widetilde{r}_t(v^m, \widetilde{b}(v^m)) \overset{\text{(a)}}{\geq} r(v^m, \widetilde{b}(v^m)) - w_t^m \overset{\text{(b)}}{\geq} r(v^m, b^k) - w_t^m \overset{\text{(c)}}{\geq} \widetilde{r}_t(v^m, b^k) - 2w_t^m,$$

where (a) and (c) follows from Lemma 3.4, and (b) holds by the definition of $\widetilde{b}(v)$. Thus, $\widetilde{b}(v^m)$ is also not eliminated by Line 9 for all t and m with probability at least $1 - \delta$, which finishes the proof.