

Bidder Subset Selection Problem in Auction Design *

Xiaohui Bei[†] Nick Gravin[‡] Pinyan Lu[†] Zhihao Gavin Tang[†]

Abstract

Motivated by practical concerns in the online advertising industry, we study a bidder subset selection problem in single-item auctions. In this problem, a large pool of candidate bidders have independent values sampled from known prior distributions. The seller needs to pick a subset of bidders and run a given auction format on the selected subset to maximize her expected revenue. We propose two frameworks for the subset restrictions: (i) capacity constraint on the set of selected bidders; and (ii) incurred costs for the bidders invited to the auction. For the second-price auction with anonymous reserve (SPA-AR), we give constant approximation polynomial time algorithms in both frameworks (in the latter framework under mild assumptions about the market). Our results are in stark contrast to the previous work of Mehta, Nadav, Psomas, Rubinstein [NeurIPS 2020], who showed hardness of approximation for the SPA without a reserve price. We also give complimentary approximation results for other well-studied auction formats such as anonymous posted pricing and sequential posted pricing. On a technical level, we find that the revenue of SPA-AR as a set function $f(S)$ of its bidders S is fractionally-subadditive but not submodular. Our bidder selection problem with invitation costs is a natural question about (approximately) answering a demand oracle for $f(\cdot)$ under a given vector of costs, a common computational assumption in the literature on combinatorial auctions.

*This work is supported by Science and Technology Innovation 2030 –“New Generation of Artificial Intelligence” Major Project No.(2018AAA0100903), Innovation Program of Shanghai Municipal Education Commission, Program for Innovative Research Team of Shanghai University of Finance and Economics (IRTSHUFE) and the Fundamental Research Funds for the Central Universities. Zhihao Gavin Tang is supported by NSFC grant 61902233. Nick Gravin is supported by NSFC grant 62150610500.

[†]Nanyang Technological University, Email: xhbei@ntu.edu.sg

[‡]ITCS, Shanghai University of Finance and Economics, Email: [{nikolai, tang.zhihao, lu.pinyan}@mail.shufe.edu.cn"> {nikolai, tang.zhihao, lu.pinyan}@mail.shufe.edu.cn](mailto)

1 Introduction

The competition among bidders and low dependency on the prior information¹ are the main contributing factors to the prevalence of auctions as selling mechanisms. An exemplary result demonstrating the importance of competition in auctions is the seminal paper by Bulow and Klemperer [BK96]. It states that a commonly used second-price single-item auction with only one additional bidder extracts more revenue than the optimal auction for n i.i.d. bidders. A more recent line of work on competition complexity initiated by Eden et al. [EFF⁺17] (see also [LP18, FFR18, BW19, CS21]) took the resource augmentation approach of [BK96] to the next level and considered the question of how many more additional bidders need to join the competition in order for a simple auction to surpass the revenue of the optimal mechanism in a variety of multi-item settings.

At a first glance it seems only beneficial for the auctioneer to have as many bidders entering the auction as possible, as the seller's revenue increases with more participants. However, this might not be the case in some real-world scenarios, as there could be inherent practical constraints on how many bidders can participate in the auction. E.g., the auction's venue may have a fixed capacity that physically prevents too many participants, or there could be time-scheduling conflicts for the bidders to simultaneously participate in the auction. In addition, the seller may incur costs for inviting and accommodating bidders in the auction, which need to be factored in when computing auctioneer's profit.

The above concerns might appear to be moot for online and digital auctions conducted on the Internet, as the online communication is easy, fast, and cheap. But this is in fact far from truth, as we demonstrate with a concrete example of real-time bidding from the online advertising industry below. Note that online advertising is a prime example of auction design applied to modern economy and serves as a major source of revenue for many tech companies.

Real-time bidding. Most online ads are selected and displayed through *real-time bidding (RTB)* – each ad slot (an impression about the advertising opportunity) is sold in real-time by an automated auction process to a set of candidate advertisers. One might think that given the online nature of the RTB auctions, the seller could invite and serve as many advertisers in the auction as possible. However, the issues we outlined above are even more pronounced in the RTB context.

- Many online advertising platforms already face (or are anticipating to face in a near future) a problem of too many advertisers participating in each online auction. It is important to keep in mind that the entire process from initiating the auction to displaying the ad on the auctioned slot takes place in just under a few *milliseconds*. Thus it is crucial for the platform to keep the communication lag and all necessary computations under a strict time limit, which is hard to achieve with a large number of bidders. The current practice is to keep this number under a threshold m (for example, $m = 20$). That is, if there are more than m bidders, the platform would need to select a subset of m bidders first before running the auction.
- Secondly, in recent years more and more advertisers are delegating their businesses to professional digital marketing agencies (DMAs), who will interact with the advertising platform and perform bidding strategies on behalf of multiple advertisers. From the advertising platform's point of view, such practices reduce the number of participants and make bidder capacity less of an issue. However, the platform then needs to spend additional effort for inviting and maintaining business relations with these agencies, which would translate into costs that need to be subtracted from the auction revenue when computing the final profit of the platform.

In all the examples above, the auctioneer is facing a *bidder selection* problem. That is, how to select an optimal subset of bidders, subject to a capacity constraint or with costs, such that the expected revenue from running an auction with these bidders is maximized. This is the topic of our paper. More concretely, we focus on the setting of selling a single item to multiple bidders with independent values and consider a few concrete auction formats used in practice.

1.1 Our Contributions We formulate a new theoretical framework for auction design with bidder selections. More formally speaking, assume that there is a large pool N of potential bidders, where each bidder $i \in N$ has a

¹It is known as the “Wilson doctrine”, named after the leading advocate for the independence of mechanism design from the precise prior knowledge about the fine details of the environment.

known prior distribution F_i . The platform selects first a set of bidders $S \subseteq N$, and then runs a specific auction format for the selected bidders. We mainly focus on the second price auction with an anonymous reserve price, which is perhaps the most relevant format in the context of online advertising. We also consider other types of auction formats, such as anonymous pricing, sequential posted pricing, and Myerson auction. We consider the following two bidder selection models.

Bidder selection with capacity constraints. In the first model, the platform needs to select a subset S of no more than $|S| \leq m$ bidders and then run a fixed auction format to maximize its expected revenue. The auction format we are mostly interested in is the second price auction with an anonymous reserve price r (the price can be optimized for the set S).

A closely related question was studied before by Mehta et al. [MNPR20]. They consider the subset selection problem for the second price auction (i.e., the reserve price $r = 0$) and show that it is NP-hard to approximate within a constant ratio. Note that no reserve price requirement is most natural for another problem of *social welfare* maximization (also known as k -MAX problem), which has been shown first by Chen et al. [CHL⁺16] to admit a PTAS (see also [MNPR20, SS21]). For the revenue objective, there is no practical restriction for the platform not to use a reserve price. We show that a subset of m bidders and a reserve price r can be found in polynomial time, so that the second price auction with reserve r achieves constant approximation to the optimum solution. This is in stark contrast to [MNPR20] whose results indicate hardness of approximation of an apparently easier optimization problem, where the reserve price r is set to be 0. Our result easily follows from (i) an observation that the revenue of the second price auction always lies within a constant factor of the optimal revenue of the anonymous pricing; and (ii) that it is easy to compute the subset of agents that maximizes revenue of the anonymous pricing mechanism.

In addition, for the sequential posted pricing mechanisms, we give a 2-approximation algorithm based on dynamic programming.

Bidder selection with invitation costs. In the second model, each bidder i has an *invitation cost* c_i that the platform needs to pay in order to include this bidder in the auction. The optimization goal of the platform is to select a subset $S \subseteq N$ of bidders with the maximum *profit margin*, i.e., the expected revenue of the auction ran on the bidder set S minus its total invitation cost $\sum_{i \in S} c_i$.

In this model, we find that the revenue of the the second price auction with anonymous reserve behaves as a fractionally subadditive function (a more general class of functions than submodular) over different subsets of bidders. The respective optimization problem with invitation costs becomes equivalent to giving an approximate solution to a demand oracle — a commonly used theoretical abstraction in the literature on combinatorial auctions. We obtain a $O(1 + \frac{1}{\delta})$ -approximation for the difference objective under a mild assumption that the market is at least δ -profitable, i.e., the revenue generated by the optimal solution is at least $(1 + \delta)$ fraction of the cost of the optimal solution. The latter assumption certainly holds for the online advertising industry, usually with a rather large value of δ . Our $O(1 + \frac{1}{\delta})$ -approximation result is much stronger than the typical bi-criteria approximation guarantees obtained for submodular maximization under costs² (see our discussion in the related works section).

We also obtain a 2-approximation result for the anonymous posted pricing mechanism. For the family of sequential posted pricing mechanisms, interestingly, our problem is closely related to the well studied pandora box problem.

	AP	AR	SPP	Myerson
Capacity Constraints	1	1.64 Theorem 3.2	2 Theorem 3.3	EPTAS [SS21, MNPR20]
Invitation Costs	2 Theorem 4.1	$O(1/\delta)$ Theorem 4.3	-	-

Table 1: Summary of Results: AP denotes the anonymous pricing mechanism. AR denotes the second price auction with anonymous reserve. SPP denotes the sequential posted pricing mechanism. Myerson denotes the Myerson auction.

²Bi-criteria results give two constant approximation guarantees when compared to the optimal solution: one for the value of the submodular function on the optimal set, another one for the total cost incurred by the optimal solution.

1.2 Related Works

k -MAX Selection. A closely related setting to our work is to select k out of n random variables so that the expected maximum value is maximized. The problem has been studied multiple times under different contexts with different names by [CHL⁺16, GGM10, SS21, MNPR20]. Chen et al. [CHL⁺16] gave a PTAS for the problem. Segev and Singla [SS21] and Mehta et al. [MNPR20] improved the result to an EPTAS. As pointed out by Mehta et al. [MNPR20], the problem has applications for search engine [BCH⁺03], procurement auction [Tan92, RL07], and team selection [KR18]. In Section 2, we observe another application to Myerson’s auction with capacity constraints.

Sequential Posted Pricing with invitation costs. We study sequential posted pricing with capacity constraints in Theorem 3.3. For the setting with invitation costs, due to the online nature of sequential posted pricing mechanisms, a natural implementation of the mechanism is to approach the bidders one by one and stop once the item is sold. In other words, the set of invited bidders are adaptive, rather than a pre-selected set of bidders. This setting is out of the scope of our bidder selection with invitation costs model. On the other hand, it has been formalized as the online pandora’s box problem by Esfandiari et al. [EHL19]. They reduced the problem to the prophet inequality setting and designed constant approximation algorithms.

Submodular maximization. Our bidder selection problem is related to the line of work on submodular maximization, where the objective is $\max_{S:|S|\leq m} f(S)$ (in the capacity constraint setting) or $\max_S f(S) - c(S)$ (in the cost setting, here c is a linear function) for a monotone submodular function f . In the capacity constraint setting, a greedy algorithm can achieve $1 - 1/e$ -approximation and this is the best possible [NWF78]. In the cost setting, although the objective function is still submodular, the greedy algorithm is not applicable as the objective function can have negative values. Sviridenko [SVW17] provide a $(1 - 1/e, 1)$ -approximation algorithm for this problem, i.e. the algorithm finds a set S with $f(S) - c(S) \geq (1 - 1/e) \cdot f(S^*) - c(S^*)$, where S^* is the optimal solution. Note that if $f(S^*)(1 - 1/e) < c(S^*)$, then [SVW17] does not give any approximation guarantees for the maximization problem $\max_S (f(S) - c(S))$ with costs. Furthermore, in general, there is no polynomial time algorithm with approximation ratio better than $n^{1-\epsilon}$ for the difference objective³. Subsequently, Feldman [Fel21] provided a guess-free algorithm and generalized the result to $(1 - e^{-\beta}, \beta)$ -approximation for every $\beta \in [0, 1]$. This is shown to be tight by Bodek and Feldman [BF22]. Recently, Lu et al. [LYG21], Sun et al. [SXZW22], and Bodek and Feldman [BF22] generalized the result to non-monotone submodular functions f and also designed bi-criteria approximations.

We remark that true approximation results are not achievable in the cost setting for general submodular functions. Our theorem 4.1 on anonymous pricing mechanisms studies a specific family of XOS functions (a.k.a. fractionally subadditive, is a super-class of submodular functions) that allows positive results. For XOS function, in the capacity constraint setting, the problem does not admit a polynomial-time $m^{1/2-\epsilon}$ -approximation algorithm in the value oracle model [Sin10], but has a 2-approximation in the demand oracle model [BDO12].

2 Preliminaries

We consider a single-item auction where the seller has one item to sell to a set of bidders. Let N denote the set of all potential bidders with $|N| = n$. Each bidder $i \in N$ has a valuation v_i for the item which is drawn from a prior distribution F_i independently across all n bidders. All prior distributions $\{F_i\}_{i \in N}$ are known to the seller. We use $f_i(p)$ and $F_i(p) = \Pr[v_i \leq p]$ to denote respectively the probability and cumulative density functions of each distribution F_i .

Auction with bidder selections. The auction has two phases. In the first phase, the seller needs to select a subset of bidders $S \subseteq N$ to participate in the auction. In the second phase, the auction takes a vector of bidders’ values $\mathbf{v} = (v_i)_{i \in S}$ as the input, and decides which bidder (if any) receives the item and how much that bidder needs to pay. In the setting with capacity constraint the auctioneer can only select a set $S : |S| \leq m$ with at most m bidders. In the setting with invitation costs, there is a known vector of costs $\mathbf{c} = (c_i)_{i \in N}$ and no restrictions on the subset S .

2.1 Auctions We will investigate the following types of auctions in this paper. Below, the set $S \subseteq N$ indicates the set of selected bidders participating in the respective auction format.

³There is an easy APX-hardness reduction for coverage valuations with linear costs to the max-independent set in a graph G on $[n]$ vertices, which is a folklore result for the unconstrained submodular maximization.

Anonymous Pricing (\mathbf{AP}_r): the seller puts up a take-it-or-leave price r to all bidders, and the item is sold to any bidder whose value for the item is no less than r . Let $\mathbf{AP}_r(S)$ denote the expected revenue of the auction with the reserve price r and bidders' set S . Then we have

$$\mathbf{AP}_r(S) = r \cdot [1 - \prod_{i \in S} F_i(r)].$$

We also use $\mathbf{AP}(S) = \max_r \mathbf{AP}_r(S)$ to denote the best revenue over all possible choices of r .

Second Price Auction with Anonymous Reserve (\mathbf{AR}_r): the seller sets a reserve price $r \geq 0$, and all bidders with values below this price are discarded; the item is sold to the highest of the remaining bidders at a price equal to the second highest agent's value or the reserve price if none other remain. We use $\mathbf{AR}_r(S)$ to denote the expected revenue of the auction with reserve price r and bidders set S , and use $\mathbf{AR}(S) = \max_r \mathbf{AR}_r(S)$ to denote the best revenue over all possible choices of r .

When the reserve price is 0, this is the classic second-price auction. We denote its expected revenue as $\mathbf{SPA}(S)$.

Sequential Posted Pricing (SPP): it is defined by the bidder set S , price vector $\mathbf{p} = (p_i)_{i \in S}$, and an order/permutation $\pi \in \Pi : S \mapsto [|S|]$ where $\pi(i)$ is the place of bidder i in the arrival sequence for each $i \in S$; the bidders arrive one by one in the order π (i.e., in the sequence $(\pi^{-1}(j))_{j \in [|S|]}$), and the item is sold to the first bidder i with $v_i \geq p_i$.

We have three notations for the expected revenue of SPP: (a) $\mathbf{SPP}(S, \mathbf{p}, \pi)$ for given prices \mathbf{p} and order π ; (b) when only the order π is fixed, $\mathbf{SPP}(S, \pi) = \max_{\mathbf{p}} \mathbf{SPP}(S, \mathbf{p}, \pi)$ is the expected revenue of the auction with optimal prices; (c) when the seller is free to choose both the order and the prices, $\mathbf{SPP}(S) = \max_{\pi} \mathbf{SPP}(S, \pi)$. For the ease of notation, we sometimes also write $\mathbf{SPP}(S, \mathbf{p}, \pi)$ when $\pi : N \mapsto [n]$ is an order of all bidders in N . In this case we assume that bidders in S arrive in the order π restricted to S .

Myerson Auction (Myerson): this is auction format that maximizes expected revenue [Mye81]; it considers the virtual value function $\phi_i(v) = v - \frac{1-F_i(v)}{f_i(v)}$ for each bidder $i \in N$; the item is allocated to the bidder (if any) with the highest non-negative virtual value $\max\{\phi_i(v_i) \mid \phi_i(v_i) \geq 0, i \in S\}$, and this bidder is charged their threshold bid to win the auction. We let $\mathbf{Myerson}(S)$ to denote the expected revenue extracted from the bidder set S .

2.2 Simple Observations

CLAIM 2.1. (SECTION 3.3.3 [HAR13]) *The expected revenue of the Myerson auction $\mathbf{Myerson}(S)$ is the same as the expected welfare of the second price auction $\widehat{\mathbf{SPA}}(S)$ for the distributions $(\widehat{F}_i)_{i \in S}$, where $\widehat{F}_i : \widetilde{v}_i = \max\{\phi_i(v_i), 0\}$ with $v_i \sim F_i$.*

This claim allows us to convert the problem about revenue maximization for the Myerson auction format to the welfare maximization problem for the second-price auction, which has been already studied before [CHL⁺16, MNPR20, SS21].

COROLLARY 2.1. ([CHL⁺16, MNPR20, SS21]) *Revenue of the Myerson auction $\mathbf{Myerson}(S)$ as a set function of the subset of bidders $S \subseteq N$ is a monotone submodular function. Moreover, there is an efficient PTAS for solving $\max_{S \subseteq N: |S| \leq m} \mathbf{Myerson}(S)$.*

For the Anonymous Pricing it is convenient to fix the price r first and then analyse $\mathbf{AP}_r(S)$.

CLAIM 2.2. *For a fixed price r , $\mathbf{AP}_r(S)$ is a monotone submodular set function of $S \subseteq N$.*

Proof. As we have $\mathbf{AP}_r(S) = r \cdot [1 - \prod_{i \in S} F_i(r)]$, the function $\mathbf{AP}_r(S)$ is monotone in S and non-negative. Furthermore, the marginal contribution $\mathbf{AP}_r(S) - \mathbf{AP}_r(S \setminus \{j\}) = r \cdot (1 - F_j(r)) \cdot \prod_{i \in S \setminus \{j\}} F_i(r)$ of a bidder $j \in S$ is a decreasing function of S . Hence, $\mathbf{AP}_r(S)$ is submodular. \square

Furthermore, it is easy to compute the optimal set $S^* = \operatorname{argmax}_{S \subseteq N, |S|=m} \mathbf{AP}_r(S)$.

⁴If the distributions $(F_i)_{i \in N}$ are regular, i.e., each $\phi_i(v)$ is monotonically non decreasing in v . For irregular distributions one should use ironed virtual functions instead (see, e.g., [Har13, Section 3.3.5] for more details).

CLAIM 2.3. For a fixed price r , it is optimal to select m bidders with the highest probability of sale in $\{\Pr[v_i \geq r]\}_{i \in N}$ as the solution $S^* = \operatorname{argmax}_{S \subseteq N, |S|=m} AP_r(S)$.

Another good property of the anonymous pricing $AP(S)$ is that it is a constant approximation to the revenue of the second-price auction with anonymous reserve $AR(S)$.

CLAIM 2.4. ([JLTX20]) For any set of bidders $S \subseteq N$ we have $1.64 \cdot AP(S) \approx \frac{\pi^2}{6} AP(S) \geq AR(S) \geq AP(S)$.

3 Capacity Constraint

In this section we study the bidder selection problem with a capacity constraint. Specifically, we consider the second price auction with anonymous reserve AR and sequential posted pricing SPP formats.

3.1 Second Price Auction with Anonymous Reserve We start by analyzing the second price auction with anonymous reserve AR . Note that when we set the reserve price $r = 0$, AR_r becomes the classic second price auction SPA (without a reserve price), and our problem becomes equivalent to selecting m out of n random variables with the objective of maximizing the expected second-highest value. This question was studied in [MNPR20] with the following conclusion.

THEOREM 3.1. ([MNPR20]) Assuming the exponential time hypothesis or the planted clique hypothesis, there is no polynomial time algorithm that, given n random variables X_1, \dots, X_n , finds a subset of size m whose expected second largest value is a constant factor of the optimal.

This means that the optimization problem $\max_{S: |S| \leq m} SPA(S)$ is APX-hard. However, this result does not necessarily imply the hardness for the objective $\max_{S: |S| \leq m} AR(S)$. In fact, as we will see below, when allowing a positive reserve price, our objective can be effectively approximated.

We start by analyzing the property of the set function $AR(S)$. It is known that certain families of set functions, such as monotone submodular functions, allow for efficient approximation algorithms. As it turns out, our function $AR(S)$ is not submodular, but belongs to a more general class of *fractionally subadditive* (also known as *XOS*) functions.

CLAIM 3.1. Revenue of the second price auction with optimal anonymous reserve $AR(S)$ as a set function of the subset of bidders $S \subseteq N$ may be not submodular, but is a monotone *XOS* function. Moreover, the additive coefficients $\{A_i \geq 0\}_{i \in S}$ in the *XOS* representation of $AR(S)$ (i.e., $AR(S) = \sum_{i \in S} A_i$ and $\forall T \subseteq N \quad AR(T) \geq \sum_{i \in T \cap S} A_i$) can be defined as follows

$$(3.1) \quad A_i = \mathbf{E}_{\mathbf{v}_i} [p_i \cdot (1 - F_i(p_i))], \text{ where } p_i(\mathbf{v}_i) = \max_{j \in S - \{i\}} (v_j, r^*) \text{ is a random posted price,}$$

$$\text{and } r^* = \operatorname{argmax}_{r \in \mathbb{R}_+} AR_r(S) \text{ is the optimal reserve price in } AR(S).$$

Proof. To see that $AR(S)$ is not submodular, consider the following instance with $n = 3$ identical bidders for some small $\varepsilon > 0$:

$$\forall i \in [3] \quad F_i \stackrel{\text{def}}{=} \begin{cases} v = 1 & \text{with probability } 99/100 \\ v \sim \text{Uni}[1, 1 + \varepsilon] & \text{with probability } 1/100 \end{cases}$$

Then for any set $S \subseteq [3]$ the optimal reserve price r^* in $AR(S)$ is $r^* = 1$. The revenue for a single bidder $AR_1(\{1\}) = 1$; for two bidders $AR_1(\{1, 2\}) = AR_1(\{1, 3\}) = 1 + \frac{1}{100^2} \cdot \frac{\varepsilon}{3}$, as the payment exceeds 1 only when both bidders have values larger than 1 and the surplus (on top of 1) is equal to the revenue of SPA for two identical and uniform $\text{Uni}[0, \varepsilon]$ bidders; for three bidders $AR_1(\{1, 2, 3\}) \geq 1 + 3 \cdot \frac{99}{100^3} \cdot \frac{\varepsilon}{3}$, as exactly 2 out of 3 bidders have their values larger than 1 with probability $3 \cdot \frac{99}{100^3}$. We get that the marginal contribution of 2 with respect to the set $\{1\}$ is smaller than its marginal contribution with respect to the set $\{1, 3\}$: $AR(\{1, 2\}) - AR(\{1\}) = \frac{\varepsilon}{3 \cdot 100^2} < \frac{\varepsilon \cdot 197}{3 \cdot 100^3} \leq AR(\{1, 2, 3\}) - AR(\{1, 3\})$.

Next we show that $AR(S)$ is an *XOS* function with the respective additive coefficients $\{A_i\}_{i \in S}$ defined in (3.1). Let us also define

$$(3.2) \quad A_i(\mathbf{v}) \stackrel{\text{def}}{=} \mathbb{1} \left[v_i \geq \max_{j \in S - \{i\}} (v_j, r^*) \right] \max_{j \in S - \{i\}} (v_j, r^*), \text{ then } A_i = \mathbf{E}_{\mathbf{v}} [A_i(\mathbf{v})],$$

where r^* is the optimal reserve price in $\text{AR}(S)$. Then we have the following (we use $\max_{i \in S}(a_i)$ to denote the second largest number in a set $\{a_i\}_{i \in S}$).

$$\begin{aligned} \text{AR}_{r^*}(S) &= \mathbf{E}_{\mathbf{v}} \left[\max_{i \in S} 2(v_i, r^*) \cdot \mathbb{1} \left[\exists i \in S : v_i \geq r^* \right] \right] = \\ &= \mathbf{E}_{\mathbf{v}} \left[\sum_{i \in S} \mathbb{1} \left[v_i \geq \max_{j \in S - \{i\}}(v_j, r^*) \right] \cdot \max_{j \in S - \{i\}}(v_j, r^*) \right] = \mathbf{E}_{\mathbf{v}} \left[\sum_{i \in S} A_i(\mathbf{v}) \right] = \sum_{i \in S} A_i. \end{aligned}$$

It remains to verify that $\text{AR}(T) \geq \sum_{i \in T} A_i$ for all $T \subseteq S$. Let $r_o = \text{argmax}_r \text{AR}_r(T)$ be the optimal price for $\text{AR}(T)$ and $\bar{T} = S \setminus T$. We shall use a random price $r \sim D$ instead of r_o given by

$$r \stackrel{\text{def}}{=} \max_{j \in \bar{T}}(v_j, r^*) \text{ for a random } \mathbf{v} \sim \mathbf{F}.$$

Then

$$\begin{aligned} \text{AR}(T) = \text{AR}_{r_o}(T) &\geq \mathbf{E}_{r \sim D} [\text{AR}_r(T)] = \mathbf{E}_{r(\mathbf{v}_{\bar{T}})} \left[\mathbf{E}_{\mathbf{v}_T} \left[\sum_{i \in T} \mathbb{1} \left[v_i \geq \max_{j \in T - \{i\}}(v_j, r) \right] \cdot \max_{j \in T - \{i\}}(v_j, r) \right] \right] = \\ &= \sum_{i \in T} \mathbf{E}_{\mathbf{v}} \left[\mathbb{1} \left[v_i \geq \max_{j \in S - \{i\}}(v_j, r^*) \right] \cdot \max_{j \in S - \{i\}}(v_j, r^*) \right] = \sum_{i \in T} A_i. \end{aligned}$$

This concludes the proof of Claim 3.1. \square

From an optimization point of view, this characterization does not immediately help with our approximation objective, as it is known that general monotone XOS functions are hard to approximate within a factor better than $m^{1/2}$ in the value oracle model [Sin10]⁵. In order to obtain an efficient approximation algorithm, we take a detour and use closely related *Anonymous Pricing*.

THEOREM 3.2. *There is a polynomial time algorithm that finds a set $S^* \subset N, |S^*| = m$ and a reserve price r^* such that $\frac{\pi^2}{6} \text{AR}_{r^*}(S^*) \geq \max_{S \subset N, |S|=m} \text{AR}(S)$.*

Proof. By Claim 2.3 we can find an optimal set $S^* \in \text{argmax}_{S \subset N, |S|=m} \text{AP}(S)$ and optimal anonymous price r^* in polynomial time by trying all possible anonymous reserve prices r . By Claim 2.4 we have

$$\frac{\pi^2}{6} \text{AR}_{r^*}(S^*) \geq \frac{\pi^2}{6} \text{AP}(S^*) = \max_{S \subset N, |S|=m} \frac{\pi^2}{6} \text{AP}(S) \geq \max_{S \subset N, |S|=m} \text{AR}(S),$$

which concludes the proof of Theorem 3.2. \square

3.2 Sequential Posted Pricing In this section we focus on the sequential posted pricing mechanisms (SPP). Recall that for any subset S of bidders, $\text{SPP}(S) = \max_{\mathbf{p}, \pi} \text{SPP}(S, \mathbf{p}, \pi)$ is the optimal revenue from running a SPP with these bidders. Our objective is to find $\max_{S: |S| \leq m} \text{SPP}(S)$.

In the following we will show a 2-approximation algorithm for this objective. The algorithm is based on two observations: (1) for any fixed bidder arrival order π , and for any subset of bidders S , the optimal revenue $\text{SPP}(S, \pi) = \max_{\mathbf{p}} \text{SPP}(S, \mathbf{p}, \pi)$ is a 2-approximation to $\text{SPP}(S)$; (2) when π is fixed, $\max_{S: |S| \leq m} \text{SPP}(S, \pi)$ can be computed efficiently by a dynamic programming algorithm.

THEOREM 3.3. *There is an efficient algorithm that provides a 2-approximation to the sequential posted price with capacity constraint problem.*

⁵Although, XOS functions do allow for efficient approximation in the *demand oracle model*. But a demand oracle is not easy to implement in our setting. Section 4 has a more detailed discussion on this issue

Proof. We first fix an arbitrary arriving order $\pi_0 = (1, 2, \dots, n)$ of the bidders. In the following we will show a dynamic programming algorithm that computes $\max_{S: |S| \leq m} \text{SPP}(S, \pi_0)$. We let

$$g(i, k) = \max_{S \subseteq [i, n], |S|=k} \text{SPP}(S, \pi_0), \quad \forall 1 \leq i \leq n, \quad 1 \leq k \leq n - i + 1$$

denote the optimal revenue with respect to the order π_0 when we are allowed to select k bidders from the set $\{i, i + 1, \dots, n\}$. The function $g(\cdot, \cdot)$ satisfies the following recurrent relation

$$g(i, k) = \max \left\{ g(i + 1, k), \max_{p_i} \left\{ p_i \cdot (1 - F_i(p_i)) + g(i + 1, k - 1) \cdot F_i(p_i) \right\} \right\}.$$

Here we consider two options: first, not to make any price offer to i ; second, offer price p_i to i , in which case we get the expected revenue of $p_i \cdot (1 - F_i(p_i))$ from i , and if i does not buy the item with probability $F_i(p_i)$, we still can make up to $k - 1$ offers to the remaining $\{i + 1, \dots, n\}$ bidders. The optimal price p_i can be found by scanning through the list of all possible prices p_i .⁶ This recurrence allows us to compute each $g(i, k)$ in a backward order, and $g(1, m)$ is the solution to our objective.

The final piece of the proof is the fact that for any fixed set of bidders S and any order π_0 , we have

$$\text{SPP}(S, \pi_0) = \max_{\mathbf{p}} \text{SPP}(S, \mathbf{p}, \pi_0) \geq \frac{1}{2} \max_{\pi} \max_{\mathbf{p}} \text{SPP}(S, \mathbf{p}, \pi) = \frac{1}{2} \text{SPP}(S)$$

where the inequality is a well-known result which can be found in, e.g., [Har13, Section 4.2.2]. \square

4 Invitation Costs

In this section we consider the setting with costs. Our main focus here will be on the second price with anonymous reserve auction format. But before that, we present a 2-approximation result for the anonymous pricing mechanisms. Let $\text{APC}(S) \stackrel{\text{def}}{=} \max_r \text{AP}_r(S) - \sum_{i \in S} c_i$ denote our objective function for any set of bidders $S \subseteq N$. Let S_o be the optimal set $S_o = \text{argmax}_S \text{APC}(S)$. We show in the following subsection how to compute a set S in polynomial time so that $2\text{APC}(S) \geq \text{APC}(S_o)$.

4.1 Anonymous Pricing We consider the problem when the price r of the anonymous pricing mechanism is fixed, i.e. we want to maximize $\text{APC}_r(S) \stackrel{\text{def}}{=} \text{AP}_r(S) - \sum_{i \in S} c_i$ for a given r . Our main result is a 2-approximation polynomial-time algorithm for this objective.

With a fixed price r every bidder i has two important parameters: 1) the probability that $\Pr[v_i \geq r]$ and 2) the invitation cost c_i . Let $w_i \stackrel{\text{def}}{=} -\ln(1 - \Pr[v_i \geq r])$. Then $\text{AP}_r(S) = r \cdot \Pr[\exists i \in S, v_i \geq r] = r \cdot (1 - \exp(-\sum_{i \in S} w_i))$. Our optimization problem can be written as follows.

$$(4.1) \quad \max_{S \subseteq N} \text{APC}_r(S) = \max_{S \subseteq N} : r \cdot \left(1 - \exp\left(-\sum_{i \in S} w_i\right)\right) - \sum_{i \in S} c_i$$

A simple observation is that we only need to consider those bidders i with $\text{APC}_r(\{i\}) \geq 0$. Formally, we prove the following.

LEMMA 4.1. $\forall i \in S_o$, we have $\text{APC}_r(\{i\}) = r \cdot (1 - \exp(-w_i)) - c_i \geq 0$.

Proof. By Claim 2.2 AP_r is a submodular function. If there is a bidder $i \in S_o$ with $\text{APC}_r(\{i\}) < 0$, then $\text{APC}_r(S_o) - \text{APC}_r(S_o \setminus \{i\}) = \text{AP}_r(S_o) - \text{AP}_r(S_o \setminus \{i\}) - c_i \leq \text{AP}_r(\{i\}) - \text{AP}_r(\emptyset) - c_i = \text{APC}_r(\{i\}) - \text{APC}_r(\emptyset) < 0$. I.e., $\text{APC}_r(S_o) < \text{APC}_r(S_o \setminus \{i\})$ - a contradiction. \square

From now on, we may assume without loss of generality that every bidder $i \in [n]$ satisfies $\text{APC}_r(\{i\}) \geq 0$. We study the following natural fractional relaxation of (4.1) similar to the fractional relaxation of the knapsack problem:

$$(4.2) \quad \max_{\mathbf{x} \in [0, 1]^n} : r \cdot \left(1 - \exp\left(-\sum_{i \in [n]} w_i \cdot x_i\right)\right) - \sum_{i \in [n]} c_i \cdot x_i$$

⁶If F_i is a continuous distribution supported on an interval $[A, B]$ ($0 < A < B$) we can find an $(1 + \varepsilon)$ approximation to all $g(i, k)$ by discretizing the possible price p_i as multiples of $\frac{\varepsilon}{n}$.

Integrality Gap. We would like to remark that the integrality gap of the above optimization problem can be arbitrarily large even when there is only 1 bidder. E.g., consider the function $h(x) = 1 - e^{-w \cdot x} - (1 - \varepsilon) \cdot x$ for $w > 1$. Then when w goes to infinity, $\max\{h(0), h(1)\} = \varepsilon - e^{-w} \rightarrow \varepsilon$, while $\max_{x \in [0,1]} h(x) \geq h(1/w) = 1 - 1/e - 1/w \rightarrow 1 - 1/e$. On the positive side, we have the unbounded integrality gap only when there is a bidder i with the corresponding *non-monotone* in $x \in [0, 1]$ function $h_i(x) = r \cdot (1 - e^{-w_i \cdot x}) - c_i \cdot x$ (equivalently, $h'(1) = r \cdot e^{-w_i} \cdot w_i - c_i < 0$). We call such bidders *special*. Let $T \stackrel{\text{def}}{=} \{i \in [n] \mid r \cdot \exp(-w_i) \cdot w_i < c_i\}$ be the set of special bidders. The next lemma shows that the optimal solution to (4.1) has at most one special bidder.

LEMMA 4.2. *The optimal set of bidders satisfies that $|S_o \cap T| \leq 1$.*

Proof. Assume towards a contradiction that there are two different bidders $j, k \in S_o \cap T$. Then

$$(4.3) \quad w_j \geq \exp(-w_k) \cdot (\exp(w_j) - 1) \quad \text{or} \quad w_k \geq \exp(-w_j) \cdot (\exp(w_k) - 1),$$

as otherwise $w_j \cdot w_k < (1 - \exp(-w_j)) \cdot (1 - \exp(-w_k))$, contradicting the fact that $w_j > 1 - \exp(-w_j)$ and $w_k > 1 - \exp(-w_k)$. Without loss of generality, we assume the first inequality of (4.3) holds. Then, we have

$$\begin{aligned} \text{APC}_r(S_o \setminus j) &= r \cdot \left(1 - \exp\left(-\sum_{i \in S_o} w_i + w_j\right) \right) - \sum_{i \in S_o} c_i + c_j > r \cdot \left(1 - \exp\left(-\sum_{i \in S_o} w_i + w_j\right) \right) \\ &\quad - \sum_{i \in S_o} c_i + r \cdot \exp(-w_j) \cdot w_j \geq r \cdot \left(1 - \exp\left(-\sum_{i \in S_o} w_i + w_j\right) \right) - \sum_{i \in S_o} c_i + r \cdot \exp(-w_k) \cdot (1 - \exp(-w_j)) \\ &\geq r \cdot \left(1 - \exp\left(-\sum_{i \in S_o} w_i + w_j\right) \right) - \sum_{i \in S_o} c_i + r \cdot (1 - \exp(-w_j)) \cdot \exp\left(-\sum_{i \in S_o} w_i + w_j\right) = \text{APC}_r(S_o), \end{aligned}$$

contradicting the optimality of S_o . Here, the first inequality holds as $j \in T$ and by the definition of T ; the second inequality holds by (4.3). \square

According to the above lemma, we can enumerate all possible choices of a bidder $t \in T \cap S_o$ and ignore all other bidders in T in the fractional relaxation (4.2). This allows us to get a slight variation of (4.2) with a bounded integrality gap. Specifically, we consider the following fractional relaxation (4.4), where the extra parameter $q \in [0, 1]$ corresponds to the event of selecting the bidder t in advance and the optimization is restricted to a subset of bidders $S \stackrel{\text{def}}{=} N \setminus T$

$$(4.4) \quad \text{FR}(S, q) : \quad \max_{\mathbf{x} \in [0,1]^S} r \cdot \left(1 - q \cdot \exp\left(-\sum_{i \in S} w_i \cdot x_i\right) \right) - \sum_{i \in S} c_i \cdot x_i$$

Note that the objective function in (4.4) is concave in \mathbf{x} . By checking local maximum conditions for the optimal solution \mathbf{x}^* to $\text{FR}(S, q)$, it is easy to see that almost the same as the standard greedy algorithm for the fractional knapsack problem finds \mathbf{x}^* (see the description of the algorithm below). Similar to the fractional knapsack problem there is at most one fractional bidder in \mathbf{x}^* .

1. Sort the bidders in S in the ascending order of w_i/c_i .
2. For i from 1 to $|S|$: continuously increase x_i until $x_i = 1$ or $r \cdot q \cdot w_i \cdot \exp(-\sum_{j \leq i} w_j \cdot x_j) = c_i$.

Finally, we provide our algorithm for approximating (4.1).

1. For each $t \in T$, solve $\text{FR}([n] \setminus T, \exp(-w_t))$ greedily and denote the optimal value and solution by $\text{FR}^*(t)$ and $\mathbf{x}^*(t)$ respectively. Let $S^*(t) = \{i \mid x_i^*(t) = 1\} \cup \{t\}$.
2. Solve $\text{FR}([n] \setminus T, 1)$ greedily and denote the optimal value and solution by $\text{FR}^*(0)$ and $\mathbf{x}^*(0)$ respectively. Let $S^*(0) = \{i \mid x_i^*(0) = 1\}$.
3. Let $i^* = \text{argmax}_i \text{APC}_r(\{i\})$.

Return the best solution among $\{i^*, S^*(0)$ and $S^*(t)$ for $t \in T$.

THEOREM 4.1. *The above algorithm provides a 2-approximation to the anonymous pricing with invitation costs problem.*

Proof. For arbitrary $t \in T \cup \{0\}$, there exists at most 1 item $i(t)$ with fractional value, i.e. $x_{i(t)} \in (0, 1)$. Since $i(t) \in [n] \setminus T$, we must have

$$(4.5) \quad c_{i(t)} \leq r \cdot \exp(-w_{i(t)}) \cdot w_{i(t)} \leq r \cdot \exp(-w_{i(t)}) \cdot \frac{\exp(w_{i(t)} \cdot (1 - x_{i(t)}^*)) - 1}{1 - x_{i(t)}^*},$$

where the second inequality holds since $e^x - 1 \geq x$ for $x \geq 0$. Consequently,

$$\begin{aligned} \text{FR}^*(t) - \text{APC}_r(S^*(t)) &= r \cdot \exp\left(-\sum_{j \in S^*(t)} w_j\right) \cdot (1 - \exp(-w_{i(t)} \cdot x_{i(t)}^*)) - c_{i(t)} \cdot x_{i(t)}^* \\ &\leq r \cdot (1 - \exp(-w_{i(t)} \cdot x_{i(t)}^*)) - c_{i(t)} \cdot x_{i(t)}^* \leq r \cdot (1 - \exp(-w_{i(t)})) - c_{i(t)} = \text{APC}_r(\{i(t)\}) \leq \text{APC}_r(\{i^*\}), \end{aligned}$$

where the second inequality is equivalent to (4.5). Finally, we conclude the proof of the theorem by Lemma 4.2 and the above inequality: $\text{APC}_r(S_o) \leq \max_{t \in T \cup \{0\}} \text{FR}^*(t)$ and

$$\max_{t \in T \cup \{0\}} \text{FR}^*(t) \leq \max_{t \in T \cup \{0\}} (\text{APC}_r(S^*(t)) + \text{APC}_r(i^*)) \leq \text{APC}_r(\{i^*\}) + \max_{t \in T \cup \{0\}} \text{APC}_r(S^*(t)).$$

Which concludes the proof of the Theorem 4.1. □

For completeness, we consider the decision version of the anonymous pricing with costs and prove its NP-hardness. Given n bidders each with value distribution F_i and cost c_i , and a number R , the question is to decide whether there exists a subset S of bidders such that $\text{APC}(S) \geq R$.

THEOREM 4.2. *The decision version of anonymous pricing with costs problem is NP-hard.*

Proof. We consider the subset-sum problem and reduce it to our problem. Given an instance $I = (\{w_1, w_2, \dots, w_n\}, W)$ for the subset-sum problem, the question is to decide whether a subset of $\{w_1, \dots, w_n\}$ sums to exactly W . We construct the following instance for the anonymous pricing with invitation costs problem with n bidders. The value distribution of each bidder i is a Bernoulli random variable: $v_i = 1$ with probability $1 - e^{-w_i}$ and $v_i = 0$ otherwise. The invitation cost of bidder i is $c_i = e^{-W} \cdot w_i$. It is easy to observe that for any subset S of bidders, the optimal anonymous pricing mechanism should set the price to be 1, i.e.

$$\text{APC}(S) = \max_r \text{AP}(S) - \sum_{i \in S} c_i = \text{AP}_1(S) - \sum_{i \in S} c_i = 1 - e^{-\sum_{i \in S} w_i} - e^{-W} \cdot \sum_{i \in S} w_i, \quad \forall S \subseteq [n].$$

Let $h(x) \stackrel{\text{def}}{=} 1 - e^{-x} - e^{-W} \cdot x$. Notice that $h'(x) = e^{-x} - e^{-W}$. The function $h(x)$ achieves its maximum value of $1 - e^{-W} - e^{-W} \cdot W$ at $x = W$. Therefore, to solve the subset-sum instance I , it suffices to decide whether $\max_S g(S) \geq 1 - e^{-W} - e^{-W} \cdot W$. Due to the NP-hardness of the subset-sum problem, we conclude the proof of the theorem. □

4.2 Second Price auction with Anonymous Reserve. Here we study the problem of maximizing the revenue of the second price auction with anonymous reserve minus the invitation cost, i.e., our objective is

$$\max_{S \subseteq N} \text{ARC}(S), \quad \text{where } \text{ARC}(S) = \text{AR}(S) - \sum_{i \in S} c_i.$$

We have already seen in Section 3 that anonymous pricing $\max_r \text{AP}_r(S^*)$ gives a computationally efficient proxy with good approximation to $\text{AR}(S^*)$. This was good enough to achieve a constant approximation for the setting with capacity constraint, but the difference objective makes the problem much more challenging. Indeed, for a general submodular function f (and our function f belongs to a more general class of XOS functions) the best we can hope for is a bi-criteria approximation of $\text{AR}(S)$ and $c(S) = \sum_{i \in S} c_i$. The difficulty of approximating the difference objective becomes quite apparent in the regime where $\text{AR}(S)$ is only slightly larger than $\sum_{i \in S} c_i$, in which case any approximation $f(S^*)$ of $\text{AR}(S^*)$ must be accurate up to $1 + \varepsilon$ factor. Thus the naive substitution of the function $\text{AR}(S^*)$ with a constant approximation $\text{AP}(S^*)$ does not work in the cost setting. Still there is a hope that $\text{APC}(S) = \text{AP}(S) - \sum_{i \in S} c_i$ might be useful in approximating $\text{ARC}(S^*)$ for a different set $S \neq S^*$.

On the positive side, we give the following approximation result parameterised by δ when the optimal solution $\text{AR}(S^*)$ has a δ -surplus over the cost $c(S^*)$, i.e., when $\text{AR}(S^*) \geq (1 + \delta) \sum_{i \in S^*} c_i$.

THEOREM 4.3. *Let S^* be the optimum solution to $\max_S \text{ARC}(S) = \max_S (\text{AR}(S) - \sum_{i \in S} c_i)$. Suppose S^* satisfies a δ -surplus condition $\text{AR}(S^*) \geq (1 + \delta) \sum_{i \in S^*} c_i$. Then there is a set of bidders S with $O\left(\frac{1+\delta}{\delta}\right) \text{APC}(S) \geq \text{ARC}(S^*)$.*

Proof. We use the probabilistic method to show existence of such set S . The idea is to randomly partition set S^* into $\ell = O\left(\frac{1}{\delta}\right)$ disjoint groups and then show that the combined revenues of the best anonymous pricing for each set in the partition almost cover $\text{AR}(S^*)$. Specifically, we assign each bidder $i \in S^*$ independently at random to one of the $[\ell]$ groups and will use a carefully chosen random price for each set in the random partition to obtain the desired bound. Let $r^* = \text{argmax}_r \text{AR}_r(S^*)$ be the optimal reserve price. We have the following lower bound on the sum of revenues for anonymous pricing applied to a random partition of S^* .

$$\begin{aligned} (4.6) \quad \mathbf{E}_{\substack{R_1, \dots, R_\ell \\ \sqcup_j R_j = S^*}} \left[\sum_{j=1}^{\ell} \text{AP}(R_j) \right] &\geq \mathbf{E}_{\substack{R_1, \dots, R_\ell \\ \sqcup_j R_j = S^*}} \left[\mathbf{E}_{\mathbf{v}} \left[\sum_{j=1}^{\ell} \overbrace{\text{AP}_{r_j(\mathbf{v})}(R_j)}^{r_j(\mathbf{v}) = \max_{k \in S^* - R_j} (v_k, r^*)} \right] \right] = \\ &\mathbf{E}_{\substack{R_1, \dots, R_\ell \\ \sqcup_j R_j = S^*}} \left[\mathbf{E}_{\mathbf{v}} \left[\sum_{i \in S^*} \sum_{j=1}^{\ell} \mathbb{1}[i \in R_j] \cdot \mathbb{1}\left[v_i \geq \max_{k \in S^* - R_j} (v_k, r^*)\right] \cdot \mathbb{1}\left[i = \text{argmax}_{k \in R_j} v_k\right] \cdot r_j(\mathbf{v}) \right] \right] \geq \\ &\mathbf{E}_{\mathbf{v}} \left[\sum_{i \in S^*} \mathbb{1}\left[i = \text{argmax}_{k \in S^*} v_k, v_i \geq r^*\right] \mathbf{E}_{\substack{R_1, \dots, R_\ell \\ \sqcup_j R_j = S^*}} \left[\sum_{j=1}^{\ell} \mathbb{1}[i \in R_j] \cdot r_j(\mathbf{v}) \right] \right] \geq \\ &\mathbf{E}_{\mathbf{v}} \left[\sum_{i \in S^*} \mathbb{1}\left[i = \text{argmax}_{k \in S^*} v_k, v_i \geq r^*\right] \cdot \frac{\ell - 1}{\ell} \max_{k \in S^*} 2(v_k, r^*) \right] = \frac{\ell - 1}{\ell} \text{AR}_{r^*}(S^*), \end{aligned}$$

where to get the first inequality we simply changed the optimal price of each set R_j to a specific random price $r_j(\mathbf{v}) = \max_{k \in S^* - R_j} (v_k, r^*)$; in the first equality we used that each v_i in \mathbf{v} has the same distribution F_i as in $\text{AP}_{r_j(\mathbf{v})}$, and also charged the price $r_j(\mathbf{v})$ to the top bidder $i \in R_j$ (we assume that argmax returns a single number, say the smallest in a lexicographic ordering, in case of a tie); to get the second inequality we rearranged the order of summations and expectations and observed that only the top bidder $i \in S^*$ can be charged the price $r_j(\mathbf{v})$ (in fact, this inequality is equality when the chance that more than one bidder has the top value is 0); to get the third inequality we observe that with probability $\frac{\ell - 1}{\ell}$ the top bidder $i \in R_j$ and the 2-nd highest bidder are assigned to different groups of the partition (R_1, \dots, R_ℓ) (in this case $r_j(\mathbf{v}) = \max_{k \in S^*} 2(v_k, r^*)$ is the same as the payment in the second price auction with anonymous reserve r^* for the set of bidders S^*).

Now, we set $\ell = \lceil \frac{2(1+\delta)}{\delta} \rceil$. Then by (4.6) there must exist a partitioning of S^* into R_1, \dots, R_ℓ such that

$$(4.7) \quad \sum_{j=1}^{\ell} \text{AP}(R_j) \geq \left(1 - \frac{1}{\ell}\right) \text{AR}_{r^*}(S^*) \geq \left(1 - \frac{\delta}{2(1+\delta)}\right) \text{AP}(S^*).$$

Finally, we get the following by summing all $\text{APC}(R_j) = \text{AP}(R_j) - \sum_{i \in R_j} c_i$.

$$\begin{aligned} \sum_{j=1}^{\ell} \text{APC}(R_j) &= \sum_{j=1}^{\ell} \text{AP}(R_j) - \sum_{i \in S^*} c_i \geq \left(1 - \frac{\delta}{2(1+\delta)}\right) \text{AR}(S^*) - \sum_{i \in S^*} c_i \\ &= \frac{1}{2} \left(\text{AR}(S^*) - \sum_{i \in S^*} c_i \right) + \frac{1}{2} \cdot \frac{\text{AR}(S^*)}{1+\delta} - \frac{1}{2} \cdot \sum_{i \in S^*} c_i \geq \frac{1}{2} \left(\text{AR}(S^*) - \sum_{i \in S^*} c_i \right) = \frac{\text{ARC}(S^*)}{2}, \end{aligned}$$

where the first inequality follows from (4.7); the second inequality holds, as $\text{AR}(S^*) \geq (1+\delta) \sum_{i \in S^*} c_i$. Therefore, for one of $j \in [\ell]$ we have $\text{APC}(R_j) \geq \frac{\text{ARC}(S^*)}{2\ell} = O\left(\frac{1+\delta}{\delta}\right) \text{ARC}(S^*)$. \square

Theorem 4.3 coupled with the computationally efficient method for finding a set T with $2 \cdot \text{APC}(T) \geq \text{APC}(S)$ from Section 4.1 gives us the following.

COROLLARY 4.1. *Let S^* be the optimum set of bidders for second price auction with anonymous reserve $S^* = \text{argmax}_S \text{ARC}(S)$. Suppose S^* satisfies a δ -surplus condition $\text{AR}(S^*) \geq (1+\delta) \sum_{i \in S^*} c_i$. Then there is polynomial time algorithm that finds a set of bidders S and a price r such that anonymous pricing for set S gets a surplus $O\left(\frac{1+\delta}{\delta}\right) \text{APC}(S) \geq \text{ARC}(S^*)$.*

On the negative side, when the generated revenue $\text{AR}(S^*) = (1+\delta) \sum_{i \in S^*} c_i$ is close to the total cost, the anonymous pricing may not approximate well $\text{ARC}(S^*)$ for any subset $S \subseteq N$.

THEOREM 4.4. *There exists a set of n i.i.d. bidders, such that the optimal profit achieved by anonymous pricing mechanism is at most $O\left(\frac{1}{\sqrt{n}}\right)$ portion of the optimal profit achieved by second price auction with anonymous reserve. I.e.,*

$$\max_S \text{APC}(S) \leq O\left(\frac{1}{\sqrt{n}}\right) \cdot \max_S \text{ARC}(S)$$

Proof. We define n auxiliary functions: $G_i(v) \stackrel{\text{def}}{=} \left(1 - \frac{i+1}{v}\right)^{\frac{1}{i}} \cdot \mathbb{1}[v \in [i+1, \infty)] \forall i \in [n]$. Consider the following instance of n i.i.d. bidders with valuation distributions $F(v) \stackrel{\text{def}}{=} \max_{i \in [n]} G_i(v)$ and invitation costs of $c = 1$ per bidder. Note that the bidders are identical as they are i.i.d. and have the same cost. We slightly abuse notations and use $\text{AP}(i), \text{AR}(i)$ to denote the revenue of anonymous pricing and the revenue of second price auction with anonymous reserve for i bidders, also use $\text{APC}(i), \text{ARC}(i)$ to denote the corresponding surplus for i bidders respectively.

We first give an upper bound of $\text{AP}(i)$:

$$\text{AP}(i) = r \cdot (1 - (F(r))^i) \leq r \cdot (1 - (G_i(r))^i) \leq i + 1.$$

Consequently, we have $\max_i \text{APC}(i) = \max_i (\text{AP}(i) - i) \leq \max_i (i + 1 - i) = 1$. We calculate next the revenue of the second price auction for n bidders:

$$\text{AR}_0(n) = n \cdot \int_{v \geq 2} v \cdot (1 - F(v)) \, dF^{n-1}(v),$$

where, we interpret the second price auction as n disjoint posted price mechanisms to each of n bidders; the price is random and is equal to the maximum among the other $n - 1$ bids.

CLAIM 4.1. *For $i \in \mathbb{N}$ and $v \geq i + 1$, we have $\left(1 - \frac{1}{v} - \frac{1}{e \cdot v^{3/2}}\right)^i \geq 1 - \frac{i+1}{v}$.*

Proof. The inequality for $i = 1$ is equivalent to $\frac{1}{v} \geq \frac{1}{e \cdot v^{3/2}}$ and holds trivially for $v \geq 2$. For $i \geq 2$, consider function $h(y) \stackrel{\text{def}}{=} (1 - y)^i - \left(1 - i \cdot y + \frac{i(i-1)}{2e^2} \cdot y^2\right)$. Observe that for $y \in [0, \frac{2}{i}]$,

$$\begin{aligned} h'(y) &= -i \cdot (1 - y)^{i-1} + i - \frac{i(i-1)}{e^2} y = iy \cdot \left((1 - y) + (1 - y)^2 + \dots + (1 - y)^{i-2} - \frac{i-1}{e^2} \right) \\ &\geq iy \cdot \left((i-1) \cdot \left(1 - \frac{2}{i}\right)^{i-2} - \frac{i-1}{e^2} \right) = i(i-1)y \cdot \left(\left(1 - \frac{2}{i}\right)^{i-2} - \frac{1}{e^2} \right) \geq 0. \end{aligned}$$

Consequently $h(y) \geq h(0) = 0$ for $y \in [0, \frac{2}{i}]$. Applying the inequality to $y = \frac{1}{v} + \frac{1}{e \cdot v^{3/2}}$, we have

$$\begin{aligned} \left(1 - \frac{1}{v} - \frac{1}{e \cdot v^{3/2}}\right)^i &\geq \left(1 - i \cdot \left(\frac{1}{v} + \frac{1}{e \cdot v^{3/2}}\right) + \frac{i(i-1)}{2e^2} \cdot \left(\frac{1}{v} + \frac{1}{e \cdot v^{3/2}}\right)^2\right) \\ &\geq 1 - \frac{i}{v} - \frac{i}{e \cdot v^{3/2}} + \frac{i^2}{4e^2} \cdot \frac{1}{v^2} = 1 - \frac{i+1}{v} + \frac{1}{v} \left(\frac{i}{2e \cdot v^{1/2}} - 1\right)^2 \geq 1 - \frac{i+1}{v}. \end{aligned}$$

This concludes the proof of the Claim. \square

According to the claim, we have that

$$v \cdot (1 - F(v)) = v \cdot \left(1 - \max_{i \leq v-1} G_i(v)\right) \geq v \cdot \left(\frac{1}{v} + \frac{1}{e \cdot v^{3/2}}\right).$$

Therefore,

$$\begin{aligned} \text{AR}_0(n) &= n \cdot \int_{v \geq 2} v \cdot (1 - F(v)) dF^{n-1}(v) \geq n \cdot \int_{v \geq 2} \left(1 + \frac{1}{e \cdot v^{1/2}}\right) dF^{n-1}(v) \\ &\geq n \cdot \left(1 + \frac{1}{e} \cdot \int_{v \in [2, n]} \frac{1}{\sqrt{v}} dF^{n-1}(v)\right) = n + \frac{\sqrt{n}}{e} \cdot F^{n-1}(n) \\ &\geq n + \frac{\sqrt{n}}{e} \cdot \max_i (G_i(n))^{n-1} \geq n + \frac{\sqrt{n}}{e} \cdot \left(1 - \frac{\sqrt{n} + 1}{n}\right)^{\frac{1}{\sqrt{n}}} = n + \Omega(\sqrt{n}). \end{aligned}$$

We conclude the proof of the theorem by noticing that $\max_i \text{ARC}(i) \geq \text{ARC}(n) \geq \text{AR}_0(n) - n \geq \Omega(\sqrt{n})$, while $\max_i \text{APC}(i) = 1$. \square

5 Conclusion

In this work we proposed two theoretical models for studying a practical problem of bidder selection in various simple auction formats. We obtained good approximation results for anonymous pricing, and found that anonymous pricing serves as a good proxy for other auction formats such as second-price auction with anonymous reserve. We obtained generally good theoretical understanding for the model with capacity constraint. However, there are still a number of open questions left in the model with invitation costs: a) if there is a PTAS, or APX hardness for anonymous posted pricing; b) whether it is possible to get a constant approximation for the $\max_S \text{ARC}(S)$ without large profit margin assumption; c) whether it is possible to obtain a constant approximation for the second price auction with the welfare objective (or revenue for the Myerson auction).

There are a few future research direction that might be useful for practical applications. To name a few, i) study bi-criteria optimization trade-offs between the revenue and welfare objectives; ii) we assumed that bidders' distributions are perfectly known to the seller, which is not always true in practice. Analyse sample complexity of the subset selection problem, or study the problem in an online learning framework. iii) From an engineering view point it is useful to have advice guided by theory about various heuristics algorithms. One approach that seem reasonable for the bidder selection problem and might be interesting from a theoretical point of view is local search algorithm. How fast it converges? What approximation guarantees can we have for the local optima?

References

- [BCH⁺03] Andrei Z. Broder, David Carmel, Michael Herscovici, Aya Soffer, and Jason Y. Zien. Efficient query evaluation using a two-level retrieval process. In *CIKM*, pages 426–434. ACM, 2003.
- [BDO12] Ashwinkumar Badanidiyuru, Shahar Dobzinski, and Sigal Oren. Optimization with demand oracles. In *Proceedings of the 13th ACM conference on electronic commerce*, pages 110–127, 2012.
- [BF22] Kobi Bodek and Moran Feldman. Maximizing sums of non-monotone submodular and linear functions: Understanding the unconstrained case. *CoRR*, abs/2204.03412, 2022.
- [BK96] Jeremy Bulow and Paul Klemperer. Auctions versus negotiations. *The American Economic Review*, 86(1):180–194, 1996.
- [BW19] Hedyeh Beyhaghi and S. Matthew Weinberg. Optimal (and benchmark-optimal) competition complexity for additive buyers over independent items. In *STOC*, pages 686–696. ACM, 2019.
- [CHL⁺16] Wei Chen, Wei Hu, Fu Li, Jian Li, Yu Liu, and Pinyan Lu. Combinatorial multi-armed bandit with general reward functions. In *NIPS*, pages 1651–1659, 2016.
- [CS21] Linda Cai and Raghuvansh R. Saxena. 99% revenue with constant enhanced competition. In *EC*, pages 224–241. ACM, 2021.
- [EFF⁺17] Alon Eden, Michal Feldman, Ophir Friedler, Inbal Talgam-Cohen, and S. Matthew Weinberg. The competition complexity of auctions: A bulow-klemperer result for multi-dimensional bidders. In *Proceedings of the 2017 ACM Conference on Economics and Computation, EC '17, Cambridge, MA, USA, June 26-30, 2017*, page 343, 2017.
- [EHL19] Hossein Esfandiari, Mohammad Taghi Hajiaghayi, Brendan Lucier, and Michael Mitzenmacher. Online pandora’s boxes and bandits. In *AAAI*, pages 1885–1892. AAAI Press, 2019.
- [Fel21] Moran Feldman. Guess free maximization of submodular and linear sums. *Algorithmica*, 83(3):853–878, 2021.
- [FFR18] Michal Feldman, Ophir Friedler, and Aviad Rubinstein. 99% revenue via enhanced competition. In *EC*, pages 443–460. ACM, 2018.
- [GGM10] Ashish Goel, Sudipto Guha, and Kamesh Munagala. How to probe for an extreme value. *ACM Trans. Algorithms*, 7(1):12:1–12:20, 2010.
- [Har13] Jason D Hartline. Mechanism design and approximation. *Book draft. October*, 122(1), 2013.
- [JLT20] Yaonan Jin, Pinyan Lu, Zhihao Gavin Tang, and Tao Xiao. Tight revenue gaps among simple mechanisms. *SIAM Journal on Computing*, 49(5):927–958, 2020.
- [KR18] Jon M. Kleinberg and Maithra Raghu. Team performance with test scores. *ACM Trans. Economics and Comput.*, 6(3-4):17:1–17:26, 2018.
- [LP18] Siqi Liu and Christos-Alexandros Psomas. On the competition complexity of dynamic mechanism design. In *SODA*, pages 2008–2025. SIAM, 2018.
- [LYG21] Cheng Lu, Wenguo Yang, and Suixiang Gao. Regularized non-monotone submodular maximization. *CoRR*, abs/2103.10008, 2021.
- [MNPR20] Aranyak Mehta, Uri Nadav, Alexandros Psomas, and Aviad Rubinstein. Hitting the high notes: Subset selection for maximizing expected order statistics. In *NeurIPS*, 2020.
- [Mye81] Roger B Myerson. Optimal auction design. *Mathematics of operations research*, 6(1):58–73, 1981.
- [NWF78] George Nemhauser, Laurence Wolsey, and M. Fisher. An analysis of approximations for maximizing submodular set functions—i. *Mathematical Programming*, 14:265–294, 12 1978.
- [RL07] ZHOU Rong and Zhi-xue LU. Two-stage procurement auction with bidders of asymmetric capacity. *Systems Engineering-Theory & Practice*, 27(12):36–41, 2007.
- [Sin10] Yaron Singer. Budget feasible mechanisms. In *2010 IEEE 51st Annual Symposium on foundations of computer science*, pages 765–774. IEEE, 2010.
- [SS21] Danny Segev and Sahil Singla. Efficient approximation schemes for stochastic probing and prophet problems. In *EC*, pages 793–794. ACM, 2021.
- [SVW17] Maxim Sviridenko, Jan Vondrák, and Justin Ward. Optimal approximation for submodular and supermodular optimization with bounded curvature. *Math. Oper. Res.*, 42(4):1197–1218, 2017.
- [SXZW22] Xin Sun, Dachuan Xu, Yang Zhou, and Chenchen Wu. Maximizing modular plus non-monotone submodular functions. *CoRR*, abs/2203.07711, 2022.
- [Tan92] Guofu Tan. Entry and r & d in procurement contracting. *Journal of Economic Theory*, 58(1):41–60, 1992.