Prior-Free Auctions with Ordered Bidders

Stefano Leonardi[*]
Sapienza University of Rome
Via Ariosto 25
00185, Rome, Italy
leon@dis.uniroma1.it

Tim Roughgarden[†]
Stanford University
353 Serra Mall
Stanford, CA 94305
tim@cs.stanford.edu

ABSTRACT

Prior-free auctions are robust auctions that assume no distribution over bidders' valuations and provide worst-case (input-by-input) approximation guarantees. In contrast to previous work on this topic, we pursue good prior-free auctions with non-identical bidders.

Prior-free auctions can approximate meaningful benchmarks for non-identical bidders only when "sufficient qualitative information" about the bidder asymmetry is publicly known. We consider digital goods auctions where there is a *total ordering* of the bidders that is known to the seller, where earlier bidders are in some sense thought to have higher valuations. We use the framework of Hartline and Roughgarden (STOC '08) to define an appropriate revenue benchmark: the maximum revenue that can be obtained from a bid vector using prices that are nonincreasing in the bidder ordering and bounded above by the second-highest bid. This monotone-price benchmark is always as large as the well-known fixed-price benchmark $\mathcal{F}^{(2)}$, so designing prior-free auctions with good approximation guarantees is only harder. By design, an auction that approximates the monotone-price benchmark satisfies a very strong guarantee: it is, in particular, simultaneously near-optimal for essentially every Bayesian environment in which bidders' valuation distributions have nonincreasing monopoly prices, or in which the distribution of each bidder stochastically dominates that of the next. Of course, even if there is no distribution over bidders' valuations, such an auction still provides a quantifiable input-by-input performance guarantee.

In this paper, we design a simple prior-free auction for digital goods with ordered bidders, the Random Price Restriction (RPR) auction. We prove that its expected revenue on every bid profile $\mathbf{b}$ is $\Omega(\mathcal{M}^{(2)}(\mathbf{b})/\log^* n)$, where $\mathcal{M}^{(2)}$ denotes the monotone-price benchmark and $\log^* n$ denotes the number of times that the $\log_2$ operator can be applied to n before the result drops below a fixed constant.

[*]This work was partially supported from EU ERC research grant PAAI (Practical Approximation Algorithms)
[†]Supported in part by NSF grant CCF-1016885, an ONR PECASE Award, and an AFOSR MURI grant.

STOCÕ12, May 19–22, 2012, New York, New York, USA.

1. INTRODUCTION

Suppose you own a set of goods and want to make money by selling them. What is the best way to do it? This question is non-trivial even in *digital goods auctions*, where the seller has an unlimited supply of identical goods (like mp3s), and there are n bidders, each of whom wants only one good and has a private valuation (i.e., maximum willingness-to-pay) for it.

The question becomes easy if the seller has a prior product distribution on bidders' valuations. Since supply is unlimited and valuations are independent, the seller can optimize for each bidder separately. For a bidder i with valuation distribution F_i, the expected revenue is maximized by posting a monopoly price — that is, making a "take-it-or-leave-it" offer at a price in $\text{argmax}_p[p \cdot (1 - F_i(p))]$.

What if good prior information is expensive or impossible to acquire? What if a single auction is to be re-used several times, in settings with different or not-yet-known bidder valuations? Are there *prior-free* auctions that admit more robust, "worst-case" revenue guarantees? Particularly germane to this paper, do such auctions exist when none of the bidders are identical?

1.1 Revenue Benchmarks

Goldberg et al. [11, 12] were the first to pursue prior-free auctions, and they proposed a competitive analysis framework based on *revenue benchmarks*.[1] The idea is to define a real-valued function on inputs (i.e., bid vectors) that represents an upper bound on the maximum revenue achievable by any "reasonable" auction on each input. They proposed the *fixed-price benchmark* $\mathcal{F}^{(2)}$ for digital goods auctions, defined as the maximum revenue that can be obtained from a given bid vector by offering every bidder a common posted price that is at most the second-highest bid.

Comparing the revenue of an auction to $\mathcal{F}^{(2)}$ initially looks like an "apples vs. oranges" comparison — the auction does not know bidders' valuations but can employ arbitrary prices, while the benchmark is privy to all the private information but handicapped in the prices it can use. Nevertheless, Goldberg et al. [11] demonstrated the effectiveness of the fixed-price benchmark for meaningful competitive analysis: no auction achieves more than a $\approx .42$ fraction of $\mathcal{F}^{(2)}$

[1]Different approaches to the design and analysis of auctions with non-Bayesian sellers were recently proposed by Azar et al. [3], Chen and Micali [6], and Lopomo et al. [18].

for every bid vector, and there are interesting auctions that obtain a constant fraction of this benchmark on every input. Subsequent attempts to extend this competitive analysis framework beyond digital goods auctions are surveyed by Hartline and Karlin [13].

1.2 The Bayesian Thought Experiment

To extend the revenue benchmark approach to new objective functions and asymmetric outcome spaces, Hartline and Roughgarden [14] advocated a general framework based on a "Bayesian thought experiment". Roughly, this framework works as follows. The first step is to temporarily think of bidders' valuations as drawn i.i.d. from some valuation distribution. The second step is to characterize the collection $\mathcal{C}$ of all optimal auctions that can arise — those with maximum-possible expected objective function value with respect to some valuation distribution. For example, for revenue maximization in digital goods auctions, $\mathcal{C}$ is the set of common posted prices (bidders are i.i.d. and hence have a common monopoly price). Finally, given a bid vector $\mathbf{b}$, the benchmark is defined as the maximum objective function value obtained by an auction in $\mathcal{C}$ on the input $\mathbf{b}$. In digital goods auctions, this is the maximum revenue that can be obtained by offering every bidder a common posted price — modulo the restriction of being at most the second-highest bid, the Bayesian thought experiment automatically regenerates the $\mathcal{F}^{(2)}$ benchmark. (For technical reasons, the upper bound on prices still needs to be added to permit interesting results [11].) More importantly, all benchmarks generated by this framework are automatically well motivated: if the performance of an auction is within a constant factor of such a benchmark for every input, then in particular it is simultaneously near-optimal in every Bayesian i.i.d. environment.[2] In addition, if there is no distribution over inputs, then the auction still provides a quantifiable input-by-input guarantee.

Analogs elsewhere in theoretical computer science include worst-case regret guarantees in online decision-making (e.g., if cost vectors are drawn i.i.d. from a distribution, then the optimal action is time-invariant) and static optimality in data structure design (e.g., if searches are i.i.d., then there is some fixed optimal binary search tree). The framework in [14] and some variants of it have been successfully used to extend the reach of prior-free mechanism design to new objective functions [14] and more complex environments [7, 15, 16].

1.3 Beyond I.I.D. Bidders

The primary goal of this paper is the following.

> *To design good prior-free auctions for benchmarks derived from non-identical bidders.*

Why is this non-trivial? Let's apply the Bayesian thought experiment to a digital goods auction, now assuming that bidder i's valuation is drawn (independently) from its own distribution F_i. For fixed distributions $F_1, \ldots, F_n$, the optimal auction offers each bidder its respective monopoly price. Ranging over all choices of $F_1, \ldots, F_n$, we find that the collection $\mathcal{C}$ corresponds to the set of *all posted price vectors*.[3] Thus, for every bid vector $\mathbf{b}$, there is an auction $A_{\mathbf{b}} \in \mathcal{C}$ that uses the price vector $\mathbf{b}$ and hence obtains the full welfare $\sum_{i=1}^{n} b_i$ as revenue. There is no digital goods auction that always obtains a constant fraction of the optimal welfare [11], so the Bayesian thought experiment with non-i.i.d. bidders generates a benchmark that is far too strong for meaningful competitive analysis.

The exercise above suggests the following principle for prior-free auction design with non-identical bidders.

> *Prior-free auctions can approximate benchmarks derived from non-identical bidders only if "sufficient qualitative information" about bidder asymmetry is publicly known.*

To give an easy example, suppose there is a publicly known partition of the bidders into groups of otherwise indistinguishable bidders. We then require the Bayesian thought experiment to conform to the public information, meaning that the valuations of bidders in the same group are i.i.d. draws from a distribution. Then, the optimal auctions $\mathcal{C}$ are the price vectors that offer a common posted price to each group of bidders. The induced prior-free benchmark is the maximum revenue that can be obtained from the given bid vector using such a price vector. This is essentially the same benchmark proposed in work on attribute auctions [4, 5] that predates the benchmark framework in [14]. There are prior-free digital goods auctions with expected revenue at least a constant fraction of this benchmark when every group has at least 2 bidders (by an easy reduction to the standard setup) and when there is a constant number of groups [4, 5].

1.4 Ordered Bidders and Stochastic Dominance

What about the general case when all bidders are distinguishable? We consider digital goods auctions when there is a *total ordering* of the bidders that is known to the seller. Without loss of generality, we assume that bidders are ordered $1, 2, \ldots, n$.[4] Earlier bidders are in some sense expected to have higher valuations. This information could be derived from, for example, zip codes, eBay bidding histories, credit history, previous transactions with the seller, and so on. We emphasize that the known information is only qualitative, and is not quantitative or distributional, as is standard in Bayesian auction design.

To generate a prior-free benchmark, we consider Bayesian thought experiments that conform to the known information. As a first step, call the distributions $F_1, \ldots, F_n$ *ordered* if the corresponding monopoly prices are nonincreasing. For example, the F_i's could be uniform distributions on intervals $[0, h_i]$ with nonincreasing h_i's; exponential distributions with nondecreasing rates; Gaussian distributions with nondecreasing means; and so on. Letting $(F_1, \ldots, F_n)$ range over all ordered distributions, the corresponding collection $\mathcal{C}$ of optimal auctions is the set of *monotone* price vectors $\mathbf{p}$, where $p_1 \geq \cdots \geq p_n$. We denote the induced revenue benchmark by $\mathcal{M}^{(1)}$, the maximum revenue that can

[2] This weaker goal of good *prior-independent* auctions — where a distribution over inputs is assumed and used in the analysis of a mechanism, but not in its design — is now studied in its own right [8, 9].

[3] This fact holds even if we restrict the F_i's to be, say, uniform distributions with supports $[0, h_i]$ (and hence monopoly prices $h_i/2$).

[4] Ties between bidders can also be accommodated easily, either with cosmetic changes to the auction and analysis in this paper, or by handling groups of indistinguishable bidders separately using known techniques.

be obtained from a given bid vector from a monotone price vector. For example, for a bid vector $\mathbf{b}$ that is itself monotone, with $b_1 \geq \cdots \geq b_n$, setting $\mathbf{p} = \mathbf{b}$ shows that $\mathcal{M}^{(1)}(\mathbf{b})$ is the full welfare $\sum_{i=1}^{n} b_i$. If $b_1 \leq \cdots \leq b_n$, however, then the revenue-maximizing monotone price vector is simply a constant price — equal to the bid b_i that maximizes $j \cdot b_j$. We emphasize that the benchmark $\mathcal{M}^{(1)}(\mathbf{b})$ is defined for *all* bid vectors $\mathbf{b}$, including those that defy the semantics of the bidder ordering. Similarly, auctions that strive to approximate such a benchmark on every input are allowed to use arbitrary prices, not merely monotone ones.

By definition, an auction with revenue at least a constant fraction of $\mathcal{M}^{(1)}$ on every input is simultaneously near-optimal in every Bayesian digital goods auction with independent and ordered distributions. Being prior-free, such an auction also has a well-defined worst-case performance guarantee even when no distribution over the inputs is assumed.

A similar simultaneous approximation result holds under the standard notion of stochastic dominance. Recall that a distribution F_i stochastically dominates another F_{i+1} if $F_i(x) \leq F_{i+1}(x)$ for every $x \geq 0$. Dhangwotnotai and Hartline (personal communication, November 2011) observed that, if F_i stochastically dominates F_{i+1} for every $i = 1, 2, \ldots, n-1$, then there is a monotone price vector with expected revenue at least 50% of that of an optimal price vector. It follows that an auction with revenue at least a constant fraction of $\mathcal{M}^{(1)}$ on every input is simultaneously near-optimal in every Bayesian digital goods auction in which the distribution of each bidder stochastically dominates that of the next.

1.5 The Monotone Price Benchmark

Given a digital goods environment with ordered bidders, we define the *monotone price benchmark* $\mathcal{M}^{(2)}(\mathbf{b})$ for every bid vector $\mathbf{b}$ as the maximum revenue obtainable via a monotone price vector in which every price is at most the second-highest bid. As in the conventional model with indistinguishable bidders [11], the upper bound on prices is necessary for the existence of prior-free auctions with non-trivial approximation guarantees.[5] Indeed, since a constant price vector is monotone, $\mathcal{M}^{(2)}(\mathbf{b}) \geq \mathcal{F}^{(2)}(\mathbf{b})$ for every $\mathbf{b}$ and so designing auctions competitive with the monotone-price benchmark is at least as difficult as with the fixed-price benchmark. Simple arguments show that $\mathcal{M}^{(2)}(\mathbf{b})$ can exceed $\mathcal{F}^{(2)}(\mathbf{b})$ by as much as a $\Theta(\log n)$ factor. As far as we know, all of the auctions previously designed to be $O(1)$-competitive with $\mathcal{F}^{(2)}$ are only $\Omega(\log n)$-competitive with $\mathcal{M}^{(2)}$.

The monotone-price benchmark was previously considered, with a completely different motivation, by Aggarwal and Hartline [1]. In [1], which predates the benchmark framework in [14], $\mathcal{M}^{(2)}$ was one of three ad hoc benchmarks proposed for "knapsack auctions", where bidders have a public size and feasible solutions correspond to subsets of bidders with total size at most a publicly known budget. In particular, Aggarwal and Hartline [1] gave a digital goods auction that, for every bid vector $\mathbf{b}$, has expected revenue at least $\frac{1}{c}\mathcal{M}^{(2)}(\mathbf{b}) - O(h \log\log\log h)$, where $c > 0$ is a constant and h is the ratio between the maximum and minimum bids. Because of the additive loss term in this guarantee, it is not directly comparable to ours. The technical approach in [1] is to classify instances as one of two types, those in which most of the optimal revenue comes from large groups of bidders such that members of the same group are charged roughly the same price, and those in which most of the optimal revenue comes from a small number of bidders who are charged a small number of different prices. Aggarwal and Hartline [1] use general techniques [4, 5] to design an auction (with some additive loss) for each of these cases, and then choose one of the two auctions at random.

1.6 Our Results

Our main result is a prior-free digital goods auction with ordered bidders that satisfies the following guarantee: there is a constant $c > 0$ such that, for every input $\mathbf{b}$, the auction's expected revenue is at least $\mathcal{M}^{(2)}(\mathbf{b})/c \log^* n$. Here $\log^* n$ denotes the number of times that the $\log_2$ operator can be applied to n before the result drops below a constant (2, say). Put differently, if n is roughly equal to a tower of t 2's, then $\log^* n$ is roughly t.

Our auction, which we call the RANDOM PRICE RESTRICTION (RPR) auction, is very simple. As is standard, the auction randomly partitions the bidders into two groups. It computes an optimal monotone price vector for the "training set" of bidders, subject to a randomly restricted set of permissible prices. Finally, it applies the computed price vector to the "test set" of bidders in the natural way. To obtain an approximation guarantee of $O(\log^* n)$, the random price restriction (and the analysis) has to be executed with some care.

1.7 Future Research

Our model and results suggest several concrete open questions, the solutions to which would advance the theory of auctions with prior-free or prior-independent guarantees. There is the obvious open question of whether or not our approximation guarantee of $O(\log^* n)$ can be improved to a constant factor, presumably using a different auction. Another important next step is the study of limited-supply auctions with unit-demand and ordered bidders. Prior-free guarantees with limited supply are non-trivial even with indistinguishable bidders [7], and extending these results to ordered bidders poses an intriguing challenge. Finally, it would be interesting to pursue prior-independent guarantees in the spirit of [8, 9] in Bayesian environments with ordered or stochastically dominating distributions.

2. PRELIMINARIES

In a *digital goods auction*, there is one seller and n bidders. There is an unlimited supply of identical goods. Each bidder wants only one good, and has a private — i.e., unknown to the seller — *valuation* v_i. We study direct-revelation auctions, in which the bidders report bids $\mathbf{b}$ to the seller, and the seller then decides who wins a good and at what price.[6] For a fixed (randomized) auction, we use $X_i(\mathbf{b})$ and $P_i(\mathbf{b})$ to denote the winning probability and expected payment of

[5] An auction that always has revenue at least a constant fraction of $\mathcal{M}^{(2)}$ is still simultaneously near-optimal in every Bayesian environment with ordered or stochastically dominating distributions, with somewhat worse constant factors, provided these distributions satisfy some mild extra conditions.

[6] For the questions we ask, the "Revelation Principle" (see, e.g., Nisan [20]) ensures that there is no loss of generality by considering only direct-revelation auctions.

bidder i when the bid profile is $\mathbf{b}$. As in previous works on prior-free auction design, we consider only auctions that are individually rational — meaning $P_i(\mathbf{b}) \le v_i \cdot X_i(\mathbf{b})$ for every i and $\mathbf{b}$ — and truthful, meaning that for each bidder i and fixed bids $\mathbf{b}_{-i}$ by the other bidders, bidder i maximizes its quasi-linear utility $v_i \cdot X_i(b_i, \mathbf{b}_{-i}) - P_i(b_i, \mathbf{b}_{-i})$ by setting $b_i = v_i$. Since we consider only truthful auctions, from now on we use bids $\mathbf{b}$ and valuations $\mathbf{v}$ interchangeably.

Truthful and individually rational digital goods auctions have a nice canonical form: for every bidder i there is a (possibly randomized) function $t_i(\mathbf{v}_{-i})$ that, given the valuations $\mathbf{v}_{-i}$ of the other bidders, gives bidder i a "take-it-or-leave-it offer" at the price $t_i(\mathbf{v}_{-i})$. This means that bidder i is given a good if and only if $v_i \ge t_i(\mathbf{v}_{-i})$, in which case it is charged the price $t_i(\mathbf{v}_{-i})$. It is clear that every choice $(t_1, \ldots, t_n)$ of such functions defines a truthful, individually rational digital goods auction; conversely, every such auction is equivalent to a choice of $(t_1, \ldots, t_n)$ [11]. A special case of such an auction is a *price vector* $\mathbf{p}$, in which each t_i is the constant function $t_i(\mathbf{v}_{-i}) = p_i$. As noted in Section 1, in Bayesian settings with independent valuations, price vectors maximize expected revenue over all truthful and individually rational auctions.

The *revenue* of an auction on the valuation profile $\mathbf{v}$ is the sum of the payments collected from the winners. Let $v^{(2)}$ denote the second-highest valuation of a profile $\mathbf{v}$. The *fixed-price benchmark* $\mathcal{F}^{(2)}$ is defined, for each valuation profile $\mathbf{v}$, as the maximum revenue that can be obtained from a constant price vector whose price is at most $v^{(2)}$:

$$\mathcal{F}^{(2)}(\mathbf{v}) = \max_{p \le v^{(2)}} \left(\sum_{i \,:\, v_i \ge p} p \right).$$

Now suppose there is a known ordering on the bidders, say $1, 2, \ldots, n$. The *monotone-price benchmark* $\mathcal{M}^{(2)}$ is defined analogously to $\mathcal{F}^{(2)}$, except that non-constant monotone price vectors are also permitted:

$$\mathcal{M}^{(2)}(\mathbf{v}) = \max_{v^{(2)} \ge p_1 \ge p_2 \ge \cdots \ge p_n} \left(\sum_{i \,:\, v_i \ge p_i} p_i \right). \tag{1}$$

Clearly, $\mathcal{M}^{(2)}(\mathbf{v}) \ge \mathcal{F}^{(2)}(\mathbf{v})$ for every input $\mathbf{v}$.

We reiterate that the monotonicity and upper-bound constraints are enforced only in the computation of the benchmark $\mathcal{M}^{(2)}$. Auctions, while obviously not privy to the private valuations, can employ whatever prices they see fit. This is natural for prior-free auctions and also necessary for non-trivial results [10].

Finally, when we say that an auction is *α-competitive* with or has *approximation factor α* for a benchmark, we mean that the auction's expected revenue is at least a $1/\alpha$ fraction of the benchmark for every input $\mathbf{v}$.

3. A PRIOR-FREE $O(\log^* n)$-APPROXIMATE DIGITAL GOODS AUCTION WITH ORDERED BIDDERS

3.1 The Random Price Restriction (RPR) Auction

We propose the Random Price Restriction (RPR) auction, displayed in Figure 1. We next elaborate on the steps of the auction. In the second step, in the case where $L = 0$, we run an arbitrary digital goods auction that is $O(1)$-competitive with respect to the fixed-price benchmark $\mathcal{F}^{(2)}$. The best-known approximation factor is 3.12 [17]; there are also very simple auctions with approximation factors 4 [11] and 4.68 [2]. Intuitively, this step is meant to extract good revenue from the set of bidders with valuations almost as high as the second-highest valuation.

Input: a valuation profile $\mathbf{v}$ for a totally ordered set $N = \{1, 2, \ldots, n\}$ of bidders.

1. Choose a *level* $L \in \{0, 1, 2, \ldots, \log^* n\}$ uniformly at random.
2. If $L = 0$, run a digital goods auction on $\mathbf{v}$ that is $O(1)$-competitive with respect to the benchmark $\mathcal{F}^{(2)}$, and halt.
3. Choose a subset $A \subseteq N$ uniformly at random, and partition N into the two sets A and B.
4. Compute an optimal L-feasible price vector $\mathbf{p}^A$ for A (see details in the main text).
5. Extend $\mathbf{p}^A$ to a price vector $\mathbf{p}$ on all of N by, for each $i \in B$, setting p_i equal to p^A_j, where $j = \min\{h > i : h \in A\}$. If j is undefined, set $p_i = 0$.
6. Sell items to bidders in B only, using prices $\mathbf{p}$.

Figure 1: The auction Random Price Restriction **(RPR).**

The third step of the algorithm randomly partitions the bidders into a "training set" A and a "test set" B. Almost all prior-free auctions have this structure, with the bidders in the training set setting prices for those in the test set. For simplicity, we sell (in the sixth step) only to bidders in the test set B. An obvious optimization is to sell simultaneously to bidders in A, using the bids of B; this would improve the hidden constant in our approximation guarantee by a factor of 2.

To explain the fourth step, let $\mathbf{v}^A$ denote the valuations of the bidders in the sample A. Let $\log^{(\ell)} n$ denote the result of applying the $\log_2$ operator ℓ times to n. A price vector is *ℓ-feasible for A* if it is monotone and if every price is a power of two in the range

$$\left[\frac{\mathcal{M}^{(2)}(\mathbf{v}^A)}{(\log^{(\ell-1)} n)^3}, \frac{3\mathcal{M}^{(2)}(\mathbf{v}^A)}{(\log^{(\ell)} n)^3} \right]. \tag{2}$$

(Here and throughout the paper, $\log^{(0)} n$ is defined as n.) The optimal such price vector is the one that maximizes the revenue obtained from the bidders in A. Intuitively, the optimal ℓ-feasible price vector is meant to extract good revenue from the set of bidders with valuations in the range in (2).

The fifth step applies the prices computed in the fourth step to bidders in the test set B in the natural way, with each bidder inheriting the price computed for its "nearest neighbor" in the training set A.

The RPR auction is truthful, as each bidder faces a take-it-or-leave-it offer at a price (possibly $+\infty$) that is independent of its reported valuation. We also note that the RPR auction can be implemented in polynomial time, as both $\mathcal{M}^{(2)}(\mathbf{v}^A)$ and $\mathbf{p}^A$ can be computed efficiently using dynamic programming.

Our definition of levels is designed to optimally make use of two easy special cases. The first special case is when the valuations are guaranteed to lie in a small range. This case is easy since there are few relevant prices to compete with. The second special case is when optimal prices are much less than the optimal revenue, implying that there are many winners at the optimal prices. This case is easy because random sampling works extremely well when there is a large number of winners. By defining levels so that the range of prices in a level shrinks according to the maximum price in the level, we are able to extract a large amount of expected revenue from all levels simultaneously.

3.2 The Analysis

Our main result is a prior-free approximation guarantee for the RPR auction.

Theorem 3.1 *There is a constant $c > 0$ such that, for every valuation profile $\mathbf{v}$, the expected revenue of the* RPR *auction is at least $\mathcal{M}^{(2)}(\mathbf{v})/(c \log^* n)$.*

Recall that $\log^* n$ denotes the number of times that the $\log_2$ operator can be applied to n before the result drops below a fixed constant. (The choice of the constant doesn't matter, in that it changes the definition only by an additive term.)

When convenient, we assume that the number n of bidders is at least a sufficiently large constant — if not, the approximation ratio of the RPR auction is trivially constant.

To make our analysis as modular and transparent as possible, we factor it into three parts. In particular, we exert absolutely no effort to control the constant c of Theorem 3.1. In the first part, we define a collection of events and analyze their probabilities. The second part is more intricate, and it shows that if certain events occur, then certain portions of the monotone-price benchmark $\mathcal{M}^{(2)}(\mathbf{v})$ can be charged against analogous portions of the RPR auction's revenue. The third part stitches together our arguments into a proof of Theorem 3.1.

3.2.1 Some Events and Their Probabilities

Fix a valuation profile $\mathbf{v}$. Let $\mathcal{M}^*$ denote $\mathcal{M}^{(2)}(\mathbf{v})$. Let $\mathbf{v}^A$ denote the profile induced by a (random) training set A. Let $\mathcal{E}_A$ denote the event that

$$\mathcal{M}^{(2)}(\mathbf{v}^A) \geq \frac{1}{3} \cdot \mathcal{M}^*.$$

Lemma 3.2 *For every valuation profile $\mathbf{v}$, $\Pr[\mathcal{E}_A] \geq \frac{1}{16}$.*

Proof. Let $\mathbf{p}$ achieve the maximum in (1) for $\mathbf{v}$, with revenue $\mathcal{M}^*$. With probability 1/4, the bidders with the highest and second-highest valuations lie in A. Given this event, the conditional expected revenue from bidders in A and B under the price vector $\mathbf{p}$ is at least $\mathcal{M}^*/2$ and at most $\mathcal{M}^*/2$, respectively. By Markov's inequality, the conditional expected revenue from bidders in A under $\mathbf{p}$ is at least $\frac{1}{3}\mathcal{M}^*$ with probability at least $\frac{1}{4}$. Since the bidders with highest and second-highest valuations lie in A, the projection of $\mathbf{p}$ on the bidders in A is a feasible price vector in the determination of the benchmark $\mathcal{M}^{(2)}(\mathbf{v}^A)$; the optimal price vector for A can only have higher revenue. Summarizing, with probability at least $\frac{1}{4} \cdot \frac{1}{4} = \frac{1}{16}$, $\mathcal{M}^{(2)}(\mathbf{v}^A) \geq \mathcal{M}^*/3$. □

With $\mathbf{v}$ still fixed, consider a choice of $\ell \in \{1, 2, \ldots, \log^* n\}$. A *level-$\ell$ price* is a power of 2 that lies in the range

$$\left[\frac{\mathcal{M}^*}{3(\log^{(\ell-1)} n)^3}, \frac{3\mathcal{M}^*}{(\log^{(\ell)} n)^3} \right]. \tag{3}$$

A *winning bidder* i for a price p satisfies $v_i \geq p$. A *level-ℓ triple* (i, j, p) meets the following criteria:

(i) p is a level-ℓ price

(ii) $i < j$

(iii) the number of winning bidders in $\{i, i+1, \ldots, j-1, j\}$ is at least
$$\frac{(\log^{(\ell)} n)^2}{432 \log^* n}, \tag{4}$$
and these include bidders i and j.

Observe that the set of level-ℓ triples depends only on $\mathbf{v}$ and ℓ, and not on the random choices of the RPR auction.

Lemma 3.3 *For every profile $\mathbf{v}$ and $\ell \in \{1, 2, \ldots, \log^* n\}$, there are $O((\log^{(\ell-1)} n)^6 \log^{(\ell)} n)$ level-ℓ triples.*

Proof. Every level-ℓ price p is at least $\mathcal{M}^*/3(\log^{(\ell-1)} n)^3$; by the definition of $\mathcal{M}^*$, there are at most $3(\log^{(\ell-1)} n)^3$ winning bidders for such a price. Thus, there are at most $O((\log^{(\ell-1)} n)^3)$ choices for each of i and j. Since every level-ℓ price is a power of 2 between $\mathcal{M}^*$ and $\mathcal{M}^*/3(\log^{(\ell-1)} n)^3$, there are $O(\log^{(\ell)} n)$ choices for p. □

Call a level-ℓ triple (i, j, p) *balanced* if at least $\frac{1}{3}$ of the winning bidders in $\{i, i+1, \ldots, j-1, j\}$ for price p lie in A, and similarly in B. Let $\mathcal{E}_\ell$ denote the event that every level-ℓ triple is balanced.

Lemma 3.4 *For every valuation profile $\mathbf{v}$ with n sufficiently large, for every $\ell \in \{1, 2, \ldots, \log^* n\}$, $\Pr[\mathcal{E}_\ell] \geq \frac{31}{32}$.*

Proof. Fix a level-ℓ triple (i, j, p) and a sufficiently large constant c. By definition (see (4)), the number of winning bidders in $\{i, i+1, \ldots, j-1, j\}$ for price p is at least $c \log^{(\ell)} n$, provided n is sufficiently large. We can choose c so that standard Chernoff bounds (e.g., [19]) imply that the probability that (i, j, p) is not balanced is at most $2^{-7 \log^{(\ell)} n} = (\log^{(\ell-1)} n)^{-7}$. Combining this fact with Lemma 3.3 and a Union Bound completes the proof. □

3.2.2 The Main Argument

Fix a valuation profile $\mathbf{v}$. As before, let $\mathcal{M}^*$ denote $\mathcal{M}^{(2)}(\mathbf{v})$. For $\ell \in \{1, 2, \ldots, \log^* n\}$, a *level-$\ell$ bidder i* is one satisfying

$$v_i \in \left[\frac{\mathcal{M}^*}{(\log^{(\ell-1)} n)^3}, \frac{\mathcal{M}^*}{(\log^{(\ell)} n)^3} \right]. \tag{5}$$

Also, recall that a price vector is *ℓ-feasible for the training set A* if it is monotone and if every price is a power of two in the range (2).

Lemma 3.5 *Fix an optimal price vector for $\mathbf{v}$, with revenue $\mathcal{M}^*$. For $\ell \in \{1, 2, \ldots, \log^* n\}$, let C^*_ℓ denote the revenue contributed by level-ℓ bidders. For every $\ell \in \{1, 2, \ldots, \log^* n\}$, if $\mathcal{E}_A$ holds, then there is an ℓ-feasible price vector for the full set N of bidders with revenue at least $C^*_\ell/2$.*

PROOF. Let $\mathbf{p}^*$ be an optimal price vector for $\mathbf{v}$ and fix $\ell \in \{1, 2, \ldots, \log^* n\}$. Let a and b denote the smallest and largest, respectively, powers of 2 that lie in the range (2). Derive $\mathbf{q}$ from $\mathbf{p}^*$ as follows: if $p_i^* > b$, set $q_i = b$; if $p_i^* < a$, set $q_i = a$; otherwise, set q_i to be the largest power of 2 less than or equal to p_i^*. Since $\mathbf{p}^*$ is monotone, so is $\mathbf{q}$. Hence, $\mathbf{q}$ is ℓ-feasible.

Since $\mathcal{E}_A$ holds, $\mathcal{M}^{(2)}(v^A) \in [\frac{1}{3}\mathcal{M}^*, \mathcal{M}^*]$. Comparing (2) with (5), we see that every level-ℓ bidder has valuation between a and b. Thus, $\mathbf{q}$ extracts at least half as much revenue as $\mathbf{p}^*$ from every such bidder. The lemma follows. □

Lemma 3.6 *Fix an optimal monotone price vector for* $\mathbf{v}$*, with revenue* $\mathcal{M}^*$*. For* $\ell \in \{1, 2, \ldots, \log^* n\}$*, let* C_ℓ^* *denote the revenue contributed by level-*ℓ *bidders, and suppose that*

$$C_\ell^* \geq \frac{\mathcal{M}^*}{2\log^* n}. \tag{6}$$

If $\mathcal{E}_A$ *and* $\mathcal{E}_\ell$ *hold, then there is an* ℓ*-feasible price vector for the training set* A *of bidders with revenue at least* $C_\ell^*/12$*.*

PROOF. By Lemma 3.5, there is an ℓ-feasible price vector $\mathbf{q}$ for N that extracts revenue at least $C_\ell^*/2 \geq \mathcal{M}^*/4\log^* n$. We claim that the projection $\mathbf{q}^A$ of $\mathbf{q}$ onto the training set A satisfies the conditions of the lemma. It certainly inherits ℓ-feasibility from $\mathbf{q}$.

Going through the bidders of N from 1 to n, we greedily partition them into minimal intervals of bidders that contribute at least $\mathcal{M}^*/(24\log^* n \log^{(\ell)} n)$ revenue under the price vector $\mathbf{q}$. (For simplicity, we ignore the final leftover interval). First, since $\mathbf{q}$ is ℓ-feasible, all of its prices are at most $3\mathcal{M}^*/(\log^{(\ell)} n)^3$, and hence every interval has at least $(\log^{(\ell)} n)^2/(72\log^* n)$ winning bidders. Second, ignoring for simplicity the negligible discretization error, by minimality there are at least $6\log^{(\ell)} n$ intervals.

Call an interval *constant* if every winning bidder pays the same price under $\mathbf{q}$. Since $\mathbf{q}$ is ℓ-feasible, prices are monotone and are powers of 2 in the range in (2). It follows that prices change in $\mathbf{q}$ at most $3\log^{(\ell)} n$ times. Since there are at least $6\log^{(\ell)} n$ intervals, at least half of them are constant. Every price used by $\mathbf{q}$ lies in the range (2) and hence, since $\mathcal{E}_A$ holds by assumption, in the range (3). Since $\mathcal{E}_\ell$ holds by assumption and every interval has at least $(\log^{(\ell)} n)^2/72\log^* n$ winning bidders, A contains at least $\frac{1}{3}$ of the winning bidders (under $\mathbf{q}$) of every constant interval. Since half of the intervals are constant and every interval contributes (essentially) the same revenue under $\mathbf{q}$, the revenue from A under $\mathbf{q}^A$ is at least $\frac{1}{6}$ times that from N under $\mathbf{q}$. □

Lemma 3.7 *With the notation and assumptions of Lemma 3.6, and if also the auction* RPR *chooses* $L = \ell$*, then its revenue is at least* $C_\ell^*/48$*.*

PROOF. Let $\mathbf{q}^A$ denote the optimal ℓ-feasible prices for A computed by the RPR auction. By Lemma 3.6, the price vector $\mathbf{q}^A$ extracts revenue at least $C_\ell^*/12$ from A, which by (6) is at least $\mathcal{M}^*/(24\log^* n)$. Let $\mathbf{q}$ denote the extension of $\mathbf{q}^A$ to N defined in the fifth step of the RPR auction, and $\mathbf{q}^B$ the projection of $\mathbf{q}$ onto the test set B.

Analogously to the proof of Lemma 3.6, we greedily partition the bidders of A into minimal intervals that contribute at least $\mathcal{M}^*/(144\log^* n \log^{(\ell)} n)$ revenue under the price vector $\mathbf{q}^A$. There are at least $6\log^{(\ell)} n$ intervals, each with at least $(\log^{(\ell)} n)^2/(432\log^* n)$ winning bidders. Since $\mathbf{q}^A$ is ℓ-feasible, at least half of these intervals are constant, in the sense that all bidders are offered a common price by $\mathbf{q}^A$.

Consider a constant interval I of A, with common price q_I. Since $\mathcal{E}_A$ holds by assumption and $\mathbf{q}^A$ is ℓ-feasible, the common price q_I in an interval I of A lies in the range (3). Let i, j be the first and last winning bidders (of A) in I and consider the interval $\{i, i+1, \ldots, j-1, j\}$ of N. Then (i, j, q_I) is a level-ℓ triple. Since $\mathcal{E}_\ell$ holds by assumption, the test set B contains at least half as many bidders $h \in I$ with $v_h \geq q_I$ as does A. By its definition, $\mathbf{q}$ offers every bidder in $\{i, i+1, \ldots, j-i, j\}$ the price q_I. Thus $\mathbf{q}$ extracts at least half as much revenue from the bidders in $B \cap \{i, i+1, \ldots, j-i, j\}$ as in $A \cap \{i, i+1, \ldots, j-i, j\}$. Since half of the intervals of A are constant and every interval of A contributes (essentially) the same amount of revenue, the revenue from B under $\mathbf{q}^B$ is at least $\frac{1}{4}$ times that from A under $\mathbf{q}^A$. □

3.2.3 Putting It All Together

The gist of the final argument is: for the most important levels ℓ, the auction RPR extracts revenue $\Omega(C_\ell^*)$ when $L = \ell$. Since each value of ℓ is chosen with probability $\approx 1/\log^* n$, the approximation guarantee of $O(\log^* n)$ follows.

Proof of Theorem 3.1: Fix a valuation profile $\mathbf{v}$ and an optimal price vector $\mathbf{p}^*$ achieving revenue $\mathcal{M}^* := \mathcal{M}^{(2)}(\mathbf{v})$. For $\ell \in \{1, 2, \ldots, \log^* n\}$, recall that a level-$\ell$ bidder has valuation in the range (5), and let C_ℓ^* denote the revenue contributed by such bidders to $\mathcal{M}^*$ under $\mathbf{p}^*$. Let C_0^* and $C_\perp^*$ denote the contributions of bidders with valuations more than $\mathcal{M}^*/(\log^{\log^* n} n)^3$ and less than $\mathcal{M}^*/n^3$, respectively. Define $S = \{\ell \in \{0, 1, 2, \ldots, \log^* n\} : C_\ell^* \geq \mathcal{M}^*/2\log^* n\}$ as the set of levels that contribute significant revenue. By averaging and the fact that $\mathcal{C}_\perp^* \leq \mathcal{M}^*/n^2$,

$$\sum_{\ell \in S} C_\ell^* \geq \frac{\mathcal{M}^*}{3}.$$

Let R denote the revenue of the RPR auction on the profile $\mathbf{v}$. Conditioning on the choice of L and using that revenue is always nonnegative, we have

$$\begin{aligned} \mathbf{E}_{A,\ell}[R] &\geq \sum_{\ell \in S} \mathbf{E}_A[R \mid L = \ell] \cdot \mathbf{Pr}[L = \ell] \\ &= \frac{1}{\log^* n + 1} \sum_{\ell \in S} \mathbf{E}_A[R \mid L = \ell]\,. \end{aligned}$$

Thus, to complete the proof, we only need to show that $\mathbf{E}[R \mid L = \ell] = \Omega(C_\ell^*)$ for every $\ell \in S$.

For non-zero $\ell \in S$, since the random choices of A and ℓ are independent, and since the events $\mathcal{E}_A$ and $\mathcal{E}_\ell$ depend on the choice of A only, we can write

$$\mathbf{E}_A[R \mid L = \ell] = \mathbf{E}_A[R \mid L = \ell \wedge \mathcal{E}_A \wedge \mathcal{E}_\ell] \cdot \mathbf{Pr}[\mathcal{E}_A \wedge \mathcal{E}_\ell]\,.$$

Combining Lemmas 3.2 and 3.4 with a Union Bound, $\mathbf{Pr}[\mathcal{E}_A \wedge \mathcal{E}_\ell] \geq \frac{1}{32}$. Since $\ell \in S$, Lemma 3.7 implies that $\mathbf{E}_A[R \mid L = \ell \wedge \mathcal{E}_A \wedge \mathcal{E}_\ell] \geq C_\ell^*/48$, and hence $\mathbf{E}[R \mid L = \ell] = \Omega(C_\ell^*)$.

Finally, recall that when $L = 0$, the RPR auction runs an auction that is $O(1)$-competitive with respect to the benchmark $\mathcal{F}^{(2)}$. Thus, to prove that $\mathbf{E}[R \mid L = 0] = \Omega(C_0^*)$, we only need to show that $\mathcal{F}^{(2)}(\mathbf{v}) = \Omega(C_0^*)$. Let N_0 denote the bidders with valuation more than $\mathcal{M}^*/(\log^{\log^* n} n)^3$, which is $\Omega(\mathcal{M}^*)$ by the definition of the $\log^*$ function. If there is only one bidder i in N_0, then offering all bidders the price p_i^* proves that $\mathcal{F}^{(2)}(\mathbf{v}) \geq C_0^* = p_i^*$. (Since $\mathbf{p}^*$ determines $\mathcal{M}^{(2)}(\mathbf{v})$, p_i^* is at most the second-highest valuation of $\mathbf{v}$ and is a legitimate option for the fixed-price benchmark $\mathcal{F}^{(2)}(\mathbf{v})$.) If N_0 contains at least two bidders with $q = \Omega(\mathcal{M}^*)$ being the second-highest valuation in N_0, then offering all bidders the price q shows that $\mathcal{F}^{(2)}(\mathbf{v}) \geq 2q = \Omega(\mathcal{M}^*) = \Omega(C_0^*)$, which completes the proof. □

Acknowledgements

We thank Peerapong Dhangwotnotai, Amos Fiat, Jason Hartline, and Elias Koutsoupias for several helpful discussions. We also thank the anonymous STOC reviewers for a number of useful comments.

4. REFERENCES

[1] G. Aggarwal and J. D. Hartline. Knapsack auctions. In *Proceedings of the 17th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1083–1092, 2006.

[2] S. Alaei, A. Malekian, and A. Srinivasan. On random sampling auctions for digital goods. In *Proceedings of the 10th ACM Conference on Electronic Commerce (EC)*, pages 187–196, 2009.

[3] P. Azar, J. Chen, and S. Micali. Crowdsourced Bayesian auctions. In *Proceedings of Innovations in Theoretical Computer Science (ITCS)*, pages 34–38, 2012.

[4] M.-F. Balcan, A. Blum, J. Hartline, and Y. Mansour. Mechanism Design via Machine Learning. *Journal of Computer and System Sciences*, 74(8):1245–1270, 2008.

[5] A. Blum and J. D. Hartline. Near-optimal online auctions. In *Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1156–1163, 2005.

[6] J. Chen and S. Micali. Mechanism design with set-theoretic beliefs. In *Proceedings of the 52nd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 87–96, 2011.

[7] N. Devanur and J. D. Hartline. Limited and online supply and the Bayesian foundations of prior-free mechanism design. In *Proceedings of the 10th ACM Conference on Electronic Commerce (EC)*, pages 41–50, 2009.

[8] N. Devanur, J. D. Hartline, A. R. Karlin, and T. Nguyen. A prior-independent mechanism for profit maximization in unit-demand combinatorial auctions. In *Proceedings of 7th Workshop on Internet & Network Economics*, 2011. To appear.

[9] Peerapong Dhangwatnotai, Tim Roughgarden, and Qiqi Yan. Revenue maximization with a single sample. In *Proceedings of the 11th ACM Conference on Electronic Commerce (EC)*, pages 129–138, 2010.

[10] A. Goldberg and J. Hartline. Envy-free auctions for digital goods. In *Proceedings of the 4th ACM Conference on Electronic Commerce (EC)*, pages 29–35, 2003.

[11] A. V. Goldberg, J. D. Hartline, A. Karlin, M. Saks, and A. Wright. Competitive auctions. *Games and Economic Behavior*, 55(2):242–269, 2006.

[12] A. V. Goldberg, J. D. Hartline, and A. Wright. Competitive auctions and digital goods. Technical Report STAR-TR-99.09.01, STAR Laboratory, InterTrust Tech. Corp., Santa Clara, CA, 1999.

[13] J. D. Hartline and A. Karlin. Profit maximization in mechanism design. In N. Nisan, T. Roughgarden, É. Tardos, and V. V. Vazirani, editors, *Algorithmic Game Theory*, chapter 13, pages 331–362. Cambridge University Press, 2007.

[14] J. D. Hartline and T. Roughgarden. Optimal mechanism design and money burning. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing (STOC)*, pages 75–84, 2008.

[15] J. D. Hartline and T. Roughgarden. Simple versus optimal mechanisms. In *Proceedings of the 10th ACM Conference on Electronic Commerce (EC)*, pages 225–234, 2009.

[16] J. D. Hartline and Q. Yan. Envy, truth, and optimality. In *Proceedings of the 12th ACM Conference on Electronic Commerce (EC)*, pages 243–252, 2011.

[17] T. Ichiba and K. Iwama. Averaging techniques for competitive auctions. In *Proceedings of the Workshop on Analytic Algorithmics and Combinatorics (ANALCO)*, pages 74–81, 2010.

[18] G. Lopomo, L. Rigotti, and C. Shannon. Uncertainty in mechanism design. Submitted, 2009.

[19] R. Motwani and P. Raghavan. *Randomized Algorithms*. Cambridge University Press, 1995.

[20] N. Nisan. Introduction to mechanism design (for computer scientists). In N. Nisan, T. Roughgarden, É. Tardos, and V. Vazirani, editors, *Algorithmic Game Theory*, chapter 9, pages 209–241. Cambridge University Press, 2007.