Games and Economic Behavior 110 (2018) 139–164

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

The complexity of optimal multidimensional pricing for a unit-demand buyer [☆]

Xi Chen [a,1], Ilias Diakonikolas [b,2], Dimitris Paparas [c,*,3], Xiaorui Sun [d,4], Mihalis Yannakakis [a,5]

[a] Department of Computer Science, Columbia University, New York, NY 10027, United States of America
[b] Department of Computer Science, University of Southern California, Los Angeles, CA 90089, United States of America
[c] Department of Computer Science, University of Wisconsin-Madison, Madison, WI 53706, United States of America
[d] Simons Institute for the Theory of Computing, University of California at Berkeley, Berkeley, CA 94720, United States of America

ARTICLE INFO

Article history:
Received 4 April 2014
Available online 10 April 2018

JEL classification:
00-01
99-00

Keywords:
Deterministic mechanism design
NP-completeness
Complexity theory
Algorithms

ABSTRACT

We resolve the complexity of revenue-optimal deterministic auctions in the unit-demand single-buyer Bayesian setting, i.e., the optimal item pricing problem, when the buyer's values for the items are independent. We show that the problem of computing a revenue-optimal pricing can be solved in polynomial time for distributions of support size 2, and its decision version is NP-complete for distributions of support size 3. We also show that the problem remains NP-complete for the case of identical distributions.

1. Introduction

Consider the following natural pricing scenario: We have a set of n items for sale and a single *unit-demand* buyer, i.e., a consumer interested in obtaining at most one of the items. The goal of the seller is to set prices for the items in order to maximize her revenue by exploiting stochastic information about the buyer's preferences. More specifically, the seller is given access to a distribution $\mathcal{F}$ from which the buyer's valuations $\mathbf{v} = (v_1, \dots, v_n)$ for the items are drawn, i.e., $\mathbf{v} \sim \mathcal{F}$, and wants to assign a price p_i to each item in order to maximize her expected revenue. The buyer's utility for item $i \in [n]$ is given by $v_i - p_i$ and she will select an item with the maximum nonnegative utility or nothing if no such item exists.

☆ An extended abstract of this work appeared in the Proceedings of the Twenty-Fifth Annual ACM–SIAM Symposium on Discrete Algorithms (SODA 2014), SIAM, Philadelphia, pp. 1319–1328.

* Corresponding author.
E-mail address: paparas@cs.wisc.edu (D. Paparas).

1 Research supported by NSF grants CCF-1149257, CCF-1423100, CCF-1703925, and a Alfred P. Sloan Foundation fellowship.

2 Work done while the author was at the University of Edinburgh, supported in part by a startup from the University of Edinburgh and a SICSA PECE grant.

3 Work done while the author was at Columbia University, supported by NSF grant CCF-1017955 and by an Alexander S. Onassis Foundation Fellowship.

4 Work done while the author was at Columbia University, supported by NSF grant CCF-1149257 and a Alfred P. Sloan Foundation fellowship.

5 Research supported by NSF grants CCF-1017955, CCF-1423100 and CCF-1703925.

https://doi.org/10.1016/j.geb.2018.03.016
0899-8256/

This problem is known as the *Bayesian Unit-demand Item-Pricing Problem* (Chawla et al., 2007) which we refer to as the *item-pricing problem* below, and has received considerable attention during the past decade (more discussion on previous work can be found in Section 1.2).

The item-pricing problem is known to have tight connections with the *optimal mechanism design problem*, a central question in mathematical economics (see Manelli and Vincent, 2007 and references therein). Finding an optimal item-pricing in our setting is equivalent to finding a revenue-optimal *deterministic* mechanism. A *randomized* mechanism, on the other hand, is more general and would allow the seller to offer *lotteries*[6] over items (Briest et al., 2010; Chawla et al., 2015). Even though randomized mechanisms in general can derive strictly more revenue (as observed in Manelli and Vincent, 2007; Thanassoulis, 2004), deterministic mechanisms (such as the item-pricings we study in this paper) are more natural and simple, and indeed they are more commonly used in practice. Optimal mechanism design is well-understood in single-parameter settings (such as the case of selling a single item to multiple buyers, including the special case of $n = 1$ in the model we study in this paper) for which Myerson (1981) obtained a closed-form characterization for the optimal mechanism; in particular, Myerson showed that in the single-parameter setting the optimal deterministic mechanism can achieve as much revenue as any randomized mechanism. The multi-parameter mechanism design problem (such as the case of selling multiple items to a single buyer studied here), however, turns out to be much more challenging.

In this paper we study the item-pricing problem with a single unit-demand buyer when $\mathcal{F} = \times_{i=1}^{n} \mathcal{F}_i$ is a *product distribution* (Chawla et al., 2007, 2010; Cai and Daskalakis, 2011), i.e., the valuations of the buyer for the n items are independent random variables. We further assume that the distributions $\mathcal{F}_i$, as the input of the problem, are *discrete* (i.e., the support of each $\mathcal{F}_i$ is a finite set) and *rational* (i.e., both values in the support of each $\mathcal{F}_i$ and their corresponding probabilities are all *rational* numbers encoded in binary). Thus the input size is the number of bits needed to represent $\mathcal{F}_i$'s. We use Item-Pricing-Opt to denote the optimization problem:

Given a product distribution, find a price vector that achieves the optimal expected revenue,

and use Item-Pricing-Decision to denote its decision version[7]:

Given a product distribution and a rational $t \geq 0$, decide if the optimal revenue is at least t.

See Section 2 for formal definitions. As is the case for most optimization problems, Item-Pricing-Opt is at least as hard as its decision version since, as we show, given any price vector, one can compute the expected revenue it achieves efficiently (see Lemma 3.1).

These computational problems exhibit very rich structures. Prior to our work, even the special case when the distributions $\mathcal{F}_i$ have support size 2 was not well understood: First note that the search space is apparently at least exponential, since the support size of $\mathcal{F}$ is 2^n. What makes things more challenging is that the optimal prices are not necessarily in the support of $\mathcal{F}$ (see Cai and Daskalakis, 2011 for a simple example with two items and distributions of support size 2). So, a priori, it was not even clear whether the optimal prices can be described with polynomially many[8] bits in the input size, whether the decision problem is in NP,[9] and whether the problems can even be solved in exponential time.

1.1. Our results

We take a principled complexity-theoretic look at the item-pricing problem with independent discrete distributions. We start by showing (Theorem 1) that the decision problem Item-Pricing-Decision is in NP (and as a corollary, the optimal prices can be described with polynomially many bits). As mentioned above, the membership proof is non-trivial because the optimal prices may not lie in the support of $\mathcal{F}$. Our proof proceeds by partitioning the space of price vectors into a set of (exponentially many) cells (defined using the input distributions $\mathcal{F}_i$), so that the optimal revenue within each cell can be computed efficiently by a shortest path computation. One consequence of the analysis is that Item-Pricing-Opt has the integrality property: if all values in the supports are integer then the optimal prices are also integer (though they may not belong to the support). Another consequence of the analysis is a simple algorithm which computes an optimal pricing by generating and evaluating a sufficient set of candidate price vectors which is guaranteed to contain an optimal price vector. The algorithm runs in polynomial space and exponential time; for a constant number of items, it runs in polynomial time. These results apply also to correlated distributions.

---

[6] A lottery in the setting of a single unit-demand buyer consists of a vector $(x_1, \ldots, x_n)$ and a price p, with $x_i \geq 0$ for all $i = 1, \ldots, n$, and $\sum_i x_i \leq 1$. If it is bought, the buyer pays p and receives an item i with probability x_i and nothing with probability $1 - \sum_i x_i$. The seller can offer a set (sometimes called a menu) of lotteries and the buyer buys one that maximizes her expected utility or nothing if every lottery in the menu has a negative utility.

[7] A decision problem is a problem that poses a "yes" or "no" question. Decision problems play a central role in computational complexity theory.

[8] This means that the number of bits is bounded from above by m^c for some constant c, where m is the input size.

[9] Informally, P is the set of all decision problems that can be solved in polynomial time and NP is the set of all decision problems for which solutions can be verified in polynomial time. A problem is NP-hard if it is at least as hard as every problem in NP (shown via polynomial-time reductions) and is NP-complete if it is both in NP and NP-hard. Many natural problems from a variety of scientific fields were shown to be NP-complete (Garey and Johnson, 1990). With NP vs P being the major open problem in theoretical computer science, NP-hard problems are generally perceived as computationally intractable. We refer interested readers to Papadimitriou (1994) for a more formal treatment of complexity theory.

Table 1
Dimensions of the pricing problem.

	Our problem BUPP	Variants in the literature
Number of buyers	a single buyer	multiple buyers
Preference	unit-demand	additive (and subadditive)
Distribution	product distributions	correlated distributions
Solution concept	deterministic (i.e., item-pricing)	randomized mechanisms
Optimality	exactly optimal	approximately optimal

We then proceed to show (Theorem 2) that when each distribution $\mathcal{F}_i$ has support size at most 2, then Item-Pricing-Opt (and thus, Item-Pricing-Decision) can be solved in polynomial time. Indeed, by exploiting the underlying structure of the problem, we show that it suffices to consider a set of $O(n^2)$ price vectors (instead of 2^n) to find the optimal price vector and compute the optimal revenue in this case.[10]

Our main result is that Item-Pricing-Decision is NP-hard (and so is Item-Pricing-Opt), even for distributions $\mathcal{F}_i$'s that have support size 3 and share the same support $\{0, 1, 3\}$ (Theorem 3) or distributions that are identical but have large support (Theorem 4). This answers an open problem first posed in Chawla et al. (2007), and later asked in Cai and Daskalakis (2011), Daskalakis et al. (2014) as well, regarding the complexity of the item-pricing problem. Our result shows that the problem of many items and a single buyer is much more complex than the case of one item and many buyers, which has a simple elegant solution as shown by Myerson; the hardness result provides compelling evidence that a similar simple structure does not exist in the multi-item case, unless P = NP. In terms of proof techniques, the main difficulty of the hardness proof stems from the fact that, for a general instance of the item-pricing problem, the expected revenue is a highly complex nonlinear function of the prices. The challenge is then to construct a family of instances such that their revenues can be well-approximated by a function that has a relatively simple expression but is at the same time general enough to encode an NP-hard problem.

1.2. Previous work

A number of variants of the problem we study in this paper have been investigated in the literature. They differ in one or more of the dimensions listed in Table 1. We will focus our discussion below on the *computational* aspects of the basic problem with a *single unit-demand buyer*. We refer interested readers to the literature for the other variants, in particular we refer to Hart and Nisan (2012, 2013), Daskalakis and Weinberg (2012), Li and Yao (2013), Cai and Huang (2013), Daskalakis et al. (2014), Babaioff et al. (2014) for results and further references on the problem with a single additive (or subadditive, Rubinstein and Weinberg, 2015) buyer and to Chawla et al. (2015), Yao (2015), Cai et al. (2016), Chawla and Miller (2016) for the problem with multiple additive or unit-demand buyers. We also refer readers to the following papers for works that give sufficient conditions on $\mathcal{F}_i$'s under which an optimal mechanism can be shown to take a specific form: Pavlov (2010) gives a sufficient condition under which an optimal menu of lotteries for an additive buyer contains boundary points only, Tang and Wang (2017) obtain structural results under certain mild conditions for selling two items to an additive buyer, Giannakopoulos and Koutsoupias (2014) analyze an additive buyer with uniform i.i.d. valuations for (at most) six items, Alaei et al. (2013) characterize settings for which marginal revenue maximization is optimal, Haghpanah and Hartline (2015) provide sufficient conditions for the optimality of certain simple mechanisms. In particular, Haghpanah and Hartline (2015) give a sufficient condition for the optimality of offering a uniform pricing in the setting of a unit-demand buyer with i.i.d. continuous distributions; the condition does not hold in the hard instances constructed in our proof of Theorem 4. We summarize now previous computational results for a single unit-demand buyer.

Optimal item-pricing and lottery-pricing (randomized mechanism) when the distribution $\mathcal{F}$ is correlated

In this case the input consists of the support of $\mathcal{F}$ and the corresponding probabilities, both listed explicitly as rational numbers (e.g., $n = 2$ and the two items have values $(1, 2)$ with probability 0.8 and $(3, 2)$ with probability 0.2). Guruswami et al. (2005) and subsequently Briest (2008) showed that no polynomial-time approximation scheme (PTAS)[11] exists for the item-pricing problem, under standard complexity-theoretic assumptions in the correlated case. If the number of items is constant, a PTAS was given by Hartline and Koltun (2005). Modifications of algorithms of Guruswami et al. (2005) and Hartline and Koltun (2005) can solve in polynomial time exactly the special case of the problem for a constant number of items when all probabilities in the input correlated distribution are multiples of $1/s$ for some positive integer s encoded in unary. (Our results yield an exact polynomial-time algorithm for constant number of items and arbitrary probabilities, encoded in binary as usual.) The problem of finding an optimal lottery-pricing for an arbitrary number of items when

[10] We use asymptotic notation such as $O(\cdot)$ and $\Omega(\cdot)$ (Papadimitriou, 1994). A function $f(n)$ is $O(n^2)$ if there exist two positive constants c and n_0 such that $f(n) \le cn^2$ for all $n \ge n_0$; $f(n)$ is $\Omega(n^2)$ if there exist constants c and n_0 such that $f(n) \ge cn^2$ for all $n \ge n_0$.

[11] We say the problem has a *polynomial-time approximation scheme* (PTAS) if given any constant $\epsilon > 0$, there is an algorithm that finds an item-pricing that achieves at least a $(1 - \epsilon)$-fraction of the optimal revenue achievable by any item-pricing, with running time polynomial on the input size (though the dependency of the running time on ϵ can be arbitrary); we say an approximation scheme is *quasi-polynomial-time* if its running time is $O(n^{\log^a n})$ for some positive constant a. Moreover, it is a *fully* polynomial-time approximation scheme (FPTAS) if the running time is polynomial in both the input size and $1/\epsilon$.

$\mathcal{F}$ is correlated, on the other hand, can be solved exactly in polynomial time via linear programming (Briest et al., 2010).[12] Briest et al. (2010) further showed that the ratio of optimal expected revenues obtained by a lottery-pricing and by an item-pricing can be unbounded in instances with four items. This was shown to hold for two items by Hart and Nisan (2013).

Optimal item-pricing when $\mathcal{F}$ is a product distribution

This is the main problem we study in this paper. In Cai and Daskalakis (2011) authors obtained a polynomial-time approximation scheme for the problem when $\mathcal{F}_i$'s are Monotone Hazard Rate distributions. They also presented a quasi-polynomial-time approximation scheme when $\mathcal{F}_i$'s lie in the broader class of regular distributions. For general product distributions, Chawla et al. (2007) obtained a 3-approximation algorithm for the problem, i.e., a polynomial-time algorithm which computes a pricing whose expected revenue is within a factor 3 or less of the optimal revenue; this was subsequently improved to a 2-approximation algorithm in Chawla et al. (2010). (These results are based on a theory of sequential posted price mechanisms developed in Chawla et al., 2007, 2007, 2010, where it is shown that such mechanisms can achieve a large fraction of the optimal revenue under single-parameter settings and generalize naturally to multi-parameter settings.) The complexity of finding an (exact) optimal item-pricing was posed as an open problem in Chawla et al. (2007, 2010), which we settle in this paper (Theorems 1 and 3). Before our work, Daskalakis et al. (2012) showed that the same problem we study in this paper is SQRT-SUM-hard[13] when either the support values or the probabilities are *irrational*. We note that their reduction relies on the fact that, for certain carefully constructed instances with irrational data, it is SQRT-SUM-hard to compare the revenue of two price vectors. This has no bearing on the complexity of the problem under the standard discrete model we consider in which all numbers are rational, for which the revenue of a price vector can be computed efficiently (see Lemma 3.1).

Optimal lottery-pricing (randomized mechanism) when $\mathcal{F}$ is a product distribution

Thanassoulis (2004) first gave a simple distribution $\mathcal{F}$ (with two items whose values are drawn independently and uniformly from Briest et al., 2010; Cai and Daskalakis, 2011) for which a lottery-pricing (or randomized mechanism) can achieve a strictly higher revenue than any item-pricing (or deterministic mechanism). In contrast to Briest et al. (2010), Hart and Nisan (2013) for the correlated case where the ratio can be unbounded (for at least two items), Chawla et al. (2015) showed that the ratio of the optimal revenues achievable by a lottery-pricing and by an item-pricing can be at most 4. Such an item-pricing can also be found efficiently (Chawla et al., 2007, 2010). On the other hand, it was shown recently (Chen et al., 2015) that the problem of finding an (exact) optimal lottery-pricing when $\mathcal{F}$ is a product distribution is intractable under standard complexity-theoretic assumptions.

1.3. Organization

The rest of the paper is organized as follows. In Section 2 we define formally the problem, state our main results, and prove some basic properties. In Section 3 we show that the decision problem is in NP, and give a simple (exponential-time) algorithm. In Section 4 we give a polynomial-time algorithm for distributions with support size 2. Section 5 shows NP-hardness for the case of support size 3, and Section 6 for the case of identical distributions. Finally we conclude in Section 7.

2. Preliminaries

2.1. Problem definition and main results

In our setting there is one buyer and one seller with n items, indexed by $[n] = \{1, 2, \dots, n\}$. The buyer is interested in buying at most one item (unit demand) and her valuation of the items is drawn from a *discrete product* distribution $\mathcal{F} = \times_{i=1}^n \mathcal{F}_i$. Each distribution $\mathcal{F}_i$ is supported on a finite set $V_i = \{v_{i,1}, \dots, v_{i,|V_i|}\} \subset \mathbb{R}_+$, where we use $\mathbb{R}_+$ to denote the set of nonnegative real numbers. We use $q_{i,j} > 0$, $j \in [|V_i|]$, to denote the probability of item i having value $v_{i,j}$, with $\sum_j q_{i,j} = 1$. Let $V = \times_{i=1}^n V_i$. We use $\Pr[\mathbf{v}]$ to denote the probability of the valuation vector being $\mathbf{v} = (v_1, \dots, v_n) \in V$, i.e., the product of $q_{i,j}$'s over i, j such that $i \in [n]$ and $v_i = v_{i,j}$.

In the problem, all the n distributions, i.e., V_i and $q_{i,j}$, are given to the seller explicitly as the input. (So the input size is the number of bits needed to encode all $v_{i,j}$'s and $q_{i,j}$'s in binary.) The seller then assigns a price $p_i \geq 0$ to each item. Once the price vector $\mathbf{p} = (p_1, \dots, p_n) \in \mathbb{R}_+^n$ is fixed, the buyer draws her values $\mathbf{v} = (v_1, \dots, v_n)$ from the n distributions independently: $\mathbf{v} \in V$ with probability $\Pr[\mathbf{v}]$. The buyer's utility for each item $i \in [n]$ is given by $v_i - p_i$. Let

[12] Note that the linear program has size polynomial in the support size of $\mathcal{F}$. For the correlated case, the support of $\mathcal{F}$ is given explicitly in the input and thus, the linear program has size polynomial in the input size. For the product case, however, the support size of $\mathcal{F}$ would be 2^n even if $\mathcal{F}_i$'s have support size 2 and thus, the linear program would be exponentially large. See Chen et al. (2015) discussed at the end of the introduction for a recent hardness result on the product case of the lottery-pricing problem.

[13] That is, it can be reduced to the SQRT-SUM problem: given positive integers $a_1, \dots, a_n, b$, decide whether $\sum_i \sqrt{a_i} \geq b$. The complexity status of this problem is open, e.g. it is not known in particular whether it is in NP, nor whether it is NP-hard.

$$\mathcal{U}(\mathbf{v},\mathbf{p}) = \max_{i\in[n]} (v_i - p_i)\,.$$

If $\mathcal{U}(\mathbf{v},\mathbf{p}) \geq 0$, the buyer selects an item $i \in [n]$ that maximizes her utility $v_i - p_i$, and the revenue of the seller is p_i. If $\mathcal{U}(\mathbf{v},\mathbf{p}) < 0$, the buyer does not select any item, and the revenue of the seller is 0.

Knowing the value distributions as well as the behavior of the buyer described above, the seller's objective is to compute a price vector $\mathbf{p} \in \mathbb{R}^n_+$ that maximizes the expected revenue

$$\mathcal{R}(\mathbf{p}) = \sum_{i\in[n]} p_i \cdot \Pr\big[\text{buyer selects item } i\big].$$

We use Item-Pricing-Decision to denote the following decision problem: the input consists of n discrete distributions, with $v_{i,j}$ and $q_{i,j}$ being rational and encoded in binary, as well as a rational number $t \geq 0$ encoded in binary; the problem asks whether the supremum of $\mathcal{R}(\mathbf{p})$ over $\mathbf{p} \in \mathbb{R}^n_+$ is at least t. We use Item-Pricing-Opt to denote the optimization problem: the input consists of n discrete distributions only, and the goal is to find an optimal price vector $\mathbf{p}$ that maximizes[14] the expected revenue $\mathcal{R}(\mathbf{p})$.

We note that these two problems are not well-defined without a tie-breaking rule, i.e. a rule that specifies which item the buyer selects when there are multiple items with maximum nonnegative utility. Throughout the paper, we will use the following *maximum price*[15] tie-breaking rule (which is convenient for our arguments): when there are multiple items with the maximum nonnegative utility, the buyer selects the item with the smallest index among items with the highest price. (We note that the critical part is that an item with the highest price is selected. Selecting the item with the smallest index among them is arbitrary — and does not affect the revenue; however, we need to make such a choice unique so that it makes sense to talk about "the" item selected by the buyer in the proofs.) We show in Section 2.2 that our choice of the tie-breaking rule does not affect the supremum of the expected revenue (hence, the complexity of Item-Pricing-Decision). At the same time, there always exists a vector $\mathbf{p}$ that achieves the supremum under the maximum price tie-breaking rule and thus, Item-Pricing-Opt is well-defined, where the goal is to find a $\mathbf{p}$ that achieves the supremum.

We are now ready to state our main results. First, we show in Section 3 that Item-Pricing-Decision is in NP.

Theorem 1. Item-Pricing-Decision *is in NP.*

We also present a simple exponential-time algorithm and show that the problem can be solved in polynomial time for constant number of items.

Next, we present in Section 4 a polynomial-time algorithm for Item-Pricing-Opt when all distributions have support size at most 2.

Theorem 2. Item-Pricing-Opt *can be solved in polynomial time when all distributions have support size at most* 2.

As our main result, we resolve the complexity of Item-Pricing-Decision by showing that it is NP-hard even when all distributions have support size 3 and share the same support $\{0,1,3\}$ (Section 5), or when they are identical (Section 6).

Theorem 3. Item-Pricing-Decision *is NP-hard even when every distribution is supported on* $\{0,1,3\}$.

Theorem 4. Item-Pricing-Decision *is NP-hard even when the distributions are identical.*

It follows that the optimization problem is also NP-hard in both cases. Thus, the problems cannot be solved in polynomial time unless P = NP.

2.2. Tie-breaking rules

Formally, a tie-breaking rule is a map from the set of pairs $(\mathbf{v},\mathbf{p})$ with $\mathcal{U}(\mathbf{v},\mathbf{p}) \geq 0$ to an item k such that $v_k - p_k = \mathcal{U}(\mathbf{v},\mathbf{p})$. In this section, we show that the supremum of the expected revenue over $\mathbf{p} \in \mathbb{R}^n_+$ is invariant to tie-breaking rules (Lemma 2.1), and the supremum can always be achieved by a price vector $\mathbf{p}$ under the maximum price tie-breaking rule (Lemma 2.3). Lemma 2.1 indeed holds for all distributions, not only discrete product distributions. Lemma 2.3 holds for all bounded distributions ($\mathcal{F}$ is bounded if there exist nonnegative real numbers a_i and b_i such that $\mathbf{v} \sim \mathcal{F}$ lies in $\times_{i=1}^n [a_i, b_i]$ with probability 1).

[14] At this moment it is not even clear whether there exists a price vector that achieves the supremum. We will see in Section 2.2 that this is always the case under the maximum price tie-breaking rule, to be discussed below.

[15] It may also be called the *maximum value* tie-breaking rule, since an item with the maximum price among a set of items with the same utility must also have the maximum value.

We will need some notation. Let B be the maximum price tie-breaking rule described earlier. We will denote by $\mathcal{R}(\mathbf{p})$ the expected revenue of $\mathbf{p}$ under B, and by $\mathcal{R}(\mathbf{v},\mathbf{p})$ the seller's revenue under B when the valuation vector is $\mathbf{v} \in V$. Given a price vector $\mathbf{p}$ and a valuation vector $\mathbf{v} \in V$, we also denote by $\mathcal{T}(\mathbf{v},\mathbf{p})$ the set of items with maximum nonnegative utility (so $\mathcal{T}(\mathbf{v},\mathbf{p}) = \emptyset$ iff $\mathcal{U}(\mathbf{v},\mathbf{p}) < 0$).

We show the following:

Lemma 2.1. *The supremum of the expected revenue over* $\mathbf{p} \in \mathbb{R}^n_+$ *is invariant to tie-breaking rules.*

Proof. Let B' be a tie-breaking rule. We will use $\mathcal{R}'(\mathbf{p})$ to denote the expected revenue of $\mathbf{p}$ under B' and use $\mathcal{R}'(\mathbf{v},\mathbf{p})$ to denote the seller's revenue under B' when the valuation vector is $\mathbf{v} \in V$.

It is clear that for any $\mathbf{p} \in \mathbb{R}^n_+$ and $\mathbf{v} \in V$, we have $\mathcal{R}(\mathbf{v},\mathbf{p}) \geq \mathcal{R}'(\mathbf{v},\mathbf{p})$ since B picks an item with the highest price among those that maximize the utility. Hence, it follows that $\sup_{\mathbf{p}} \mathcal{R}(\mathbf{p}) \geq \sup_{\mathbf{p}} \mathcal{R}'(\mathbf{p})$.

On the other hand, given any price vector $\mathbf{p} \in \mathbb{R}^n_+$ and $0 < \epsilon < 1$, we consider $\mathbf{p}_\epsilon = (1-\epsilon)\mathbf{p}$. This reduces the price of each item i by ϵp_i, thus increases its utility for any valuation by the same amount; hence if two items had the same utility for $\mathbf{p}$, now the higher priced one has higher utility for $\mathbf{p}_\epsilon$. For any valuation $\mathbf{v}$, the buyer buys under B' and prices $\mathbf{p}_\epsilon$ either the same item as the one bought under B and $\mathbf{p}$ or another one that had an equal or higher price. Thus, $\mathcal{R}'(\mathbf{v},\mathbf{p}_\epsilon) \geq (1-\epsilon)\mathcal{R}(\mathbf{v},\mathbf{p})$ for every valuation $\mathbf{v}$. Therefore, $\mathcal{R}'(\mathbf{p}_\epsilon) \geq (1-\epsilon)\mathcal{R}(\mathbf{p})$, hence $\lim_{\epsilon \to 0+} \mathcal{R}'(\mathbf{p}_\epsilon) \geq \mathcal{R}(\mathbf{p})$. It follows that $\sup_{\mathbf{p}} \mathcal{R}'(\mathbf{p}) \geq \sup_{\mathbf{p}} \mathcal{R}(\mathbf{p})$, which proves the lemma. □

We will henceforth always adopt the maximum price tie-breaking rule, and use $\mathcal{R}(\mathbf{v},\mathbf{p})$ to denote the revenue of the seller with respect to this rule. One advantage of this rule is that the supremum of the expected revenue $\mathcal{R}(\mathbf{p})$ is always achievable for *bounded* distributions $\mathcal{F}$, so it makes sense to talk about whether $\mathbf{p}$ is optimal or not. In the following example we point out that this does not hold for general tie-breaking rules.

Example. Suppose there are two items with the following distributions: item 1 has value 1 with probability 1, item 2 has value 0 with probability 1/2 and value 2 with probability 1/2. Suppose that the tie-breaking rule is that, in case of tie in utility, the buyer prefers item 1 (instead of the higher-priced item). The supremum in this example is 1.5: set $p_1 = 1$ for item 1 and $p_2 = 2-\epsilon$ for item 2, for any $\epsilon > 0$. The buyer will buy item 1 with probability 1/2 (if her value for item 2 is 0) and item 2 with probability 1/2 (if her value for item 2 is 2), yielding expected revenue $1.5 - 0.5\epsilon$. However, an expected revenue of 1.5 is not achievable: if we give price 2 to item 2, then the buyer will always buy item 1 and the revenue is 1. Note that the expected revenue for this tie-breaking rule is not a continuous function of the prices. □

Let $\mathcal{F}$ be a bounded distribution such that $\mathbf{v} \sim \mathcal{F}$ lies in $P = \times_{i=1}^n [a_i, b_i]$ for some nonnegative real numbers a_i and b_i with probability 1. Before proving that the supremum is achievable under the maximum price rule for $\mathcal{F}$, we show that one may focus the search for an optimal price vector in P (instead of $\mathbb{R}^n_+$).

Lemma 2.2. *For any price vector* $\mathbf{p} \in \mathbb{R}^n_+$, *there exists a* $\mathbf{p}' \in P$ *such that* $\mathcal{R}(\mathbf{p}') \geq \mathcal{R}(\mathbf{p})$.

Proof. Let $\mathbf{p} \in \mathbb{R}^n_+$. Suppose that there exists an $i \in [n]$ such that $p_i < a_i$. Let i be such an index with the maximum a_i. Then we define a new price vector $\widetilde{\mathbf{p}}$ by setting $\widetilde{p}_j = \max\{p_j, a_j\}$ for all $j \in [n]$. Note that $\widetilde{p}_j \geq a_j$ for all j and we prove below that $\mathcal{R}(\widetilde{\mathbf{p}}) \geq \mathcal{R}(\mathbf{p})$. Consider any valuation $\mathbf{v} \in V$. If the buyer selected under prices $\mathbf{p}$ an item j with price $p_j > a_i$, then she still selects the same item under $\widetilde{\mathbf{p}}$ because we have only increased the prices of the items that had price $< a_i$. If she selected under $\mathbf{p}$ an item with price $\leq a_i$, then now she selects an item with price at least a_i, because all items have price $\widetilde{p}_j \geq a_i$ (and she will buy some item because for example item i has price a_i and thus, the utility is nonnegative). Thus, for every valuation the revenue does not decrease, hence $\mathcal{R}(\widetilde{\mathbf{p}}) \geq \mathcal{R}(\mathbf{p})$.

Next we define $\mathbf{p}'$ to be the price vector with $p'_i = \min\{\widetilde{p}_i, b_i\}$ for each $i \in [n]$. Then $\mathbf{p}' \in P$. It follows from a similar analysis that, for any valuation, the revenue does not decrease and thus $\mathcal{R}(\mathbf{p}') \geq \mathcal{R}(\widetilde{\mathbf{p}})$. This finishes the proof of the lemma. □

Now we show that the supremum can always be achieved (for bounded distributions) under the maximum price rule B using a standard continuity argument.[16]

Lemma 2.3. *When* $\mathcal{F}$ *is bounded, there exists a price vector* $\mathbf{p}^* \in P$ *such that* $\mathcal{R}(\mathbf{p}^*) = \sup_{\mathbf{p}} \mathcal{R}(\mathbf{p})$.

Proof. By the compactness of P, it suffices to show that if a sequence of vectors $\{\mathbf{p}_i\}$ approaches $\mathbf{p}$, then

$$\mathcal{R}(\mathbf{p}) \geq \lim_{i \to \infty} \mathcal{R}(\mathbf{p}_i).$$

[16] One might attempt to prove Lemma 2.3 by applying a limit argument on a theorem of Nisan (Theorem 21 in Balcan et al., 2008). However, it does not work, as the proof of the latter proceeds by first assuming the existence of an optimal price vector, which is exactly what Lemma 2.3 aims to establish.

To this end, it suffices to show that, for any valuation vector $\mathbf{v} \in V$, we have

$$\mathcal{R}(\mathbf{v},\mathbf{p}) \geq \lim_{i\to\infty} \mathcal{R}(\mathbf{v},\mathbf{p}_i). \tag{1}$$

Given any valuation vector $\mathbf{v} \in V$, it is easy to check that $\mathcal{T}(\mathbf{v},\mathbf{p}_i) \subseteq \mathcal{T}(\mathbf{v},\mathbf{p})$ when i is sufficiently large (by considering two cases: $\mathcal{U}(\mathbf{v},\mathbf{p}) < 0$ and $\mathcal{U}(\mathbf{v},\mathbf{p}) \geq 0$). (1) then follows, since $\mathcal{R}(\mathbf{v},\mathbf{p})$ is the highest price of all items in $\mathcal{T}(\mathbf{v},\mathbf{p})$ under the maximum price tie-breaking rule. □

3. Membership in NP and an algorithm

In this section we prove that Item-Pricing-Decision is in NP. We then analyze further the structure of the item-pricing problem and give a simple algorithm for computing an optimal price vector in polynomial space and exponential time; the algorithm runs in polynomial time when the number of items is a constant.

We start by showing that, for any given price vector $\mathbf{p}$, we can compute its expected revenue in polynomial time.

Lemma 3.1. *Given an instance of* Item-Pricing-Decision *(n discrete probability distributions) and a price vector* $\mathbf{p}$*, we can compute its expected revenue* $\mathcal{R}(\mathbf{p})$ *in polynomial time.*

Proof. For each item i and value s_i in its support V_i, let $V(\mathbf{p}, s_i)$ be the set of valuations $\mathbf{v} \in V$ such that $v_i = s_i$ and the buyer buys item i for valuation $\mathbf{v}$ (under the maximum price tie-breaking rule). For this to be the case, first we must obviously have $s_i \geq p_i$ (otherwise $V(\mathbf{p}, s_i) = \emptyset$). Furthermore, for each $j \neq i$, the buyer must prefer item i to j, thus we must have $v_j - p_j \leq s_i - p_i$, in case of equality we must have $p_j \leq p_i$ (and $v_j \leq s_i$), and in case of further equality $p_j = p_i$, we must have $i < j$. Let $L_j(\mathbf{p}, s_i)$ be the set of $v_j \in V_j$ that satisfy these conditions. Clearly, we can compute $L_j(\mathbf{p}, s_i)$ easily (in linear time). Then

$$V(\mathbf{p}, s_i) = L_1(\mathbf{p}, s_i) \times \cdots \times L_{i-1}(\mathbf{p}, s_i) \times \{s_i\} \times L_{i+1}(\mathbf{p}, s_i) \times \cdots \times L_n(\mathbf{p}, s_i).$$

Its probability, $\Pr(V(\mathbf{p}, s_i))$, is the product of the probabilities of the subsets $L_j(\mathbf{p}, s_i)$ for all $j \neq i$ and the probability of s_i. Let $\gamma_i(\mathbf{p})$ be the probability that the buyer selects item i; this is the sum of $\Pr(V(\mathbf{p}, s_i))$ over all $s_i \in V_i$ with $s_i \geq p_i$. Having computed $\gamma_i(\mathbf{p})$ for all $i \in [n]$, the expected revenue for price vector $\mathbf{p}$ is $\mathcal{R}(\mathbf{p}) = \sum_i \gamma_i(\mathbf{p}) \cdot p_i$. This finishes the proof of the lemma. □

We show now the NP membership.

Theorem 1. Item-Pricing-Decision *is in NP.*

Proof. We will show that there is an optimal price vector whose entries are all rational numbers with a polynomial number of bits. By Lemma 3.1 above the expected revenue of a price vector can be computed in polynomial time. As a consequence, the optimum price vector $\mathbf{p}^*$ can serve as an appropriate *yes* certificate for the decision problem Item-Pricing-Decision, and the theorem follows. To prove the existence of an optimal price vector with polynomial bit complexity, we introduce polynomially many hyperplanes which divide the space of price vectors into cells such that the optimal price vector within each cell can be found by solving a linear program. This idea of dividing the search space into cells and reducing the search problem within each cell to a linear program is not new. For example, it was used by Devanur and Kannan (2008) to study the computation of market equilibria, and by Etessami and Yannakakis (2010) for fixed points of piecewise linear functions.

We start with some notation. Given a price vector $\mathbf{p} \in \mathbb{R}_+^n$ and a valuation $\mathbf{v} \in V$, let $\mathcal{I}(\mathbf{v},\mathbf{p}) \in [n] \cup \{\text{nil}\}$ denote the item picked by the buyer under the maximum price tie-breaking rule, with $\mathcal{I}(\mathbf{v},\mathbf{p}) = \text{nil}$ if $\mathcal{U}(\mathbf{v},\mathbf{p}) < 0$. Let $a_i = \min_{v \in V_i} v$ and $b_i = \max_{v \in V_i} v$ for each i. By Lemma 2.2, it suffices to consider $P = \times_{i=1}^n [a_i, b_i]$. We now partition P into equivalence classes so that two vectors $\mathbf{p}$ and $\mathbf{p}'$ from the same class yield the same outcome for all valuations: $\mathcal{I}(\mathbf{v},\mathbf{p}) = \mathcal{I}(\mathbf{v},\mathbf{p}')$ for all $\mathbf{v} \in V$.[17]

Consider the partition of P induced by the following set of hyperplanes. For each item $i \in [n]$ and each value $s_i \in V_i$, we have a hyperplane $p_i = s_i$. For each pair of items $i, j \in [n]$ and pair of values $s_i \in V_i$ and $t_j \in V_j$, we have a hyperplane $s_i - p_i = t_j - p_j$, i.e., $p_i - p_j = s_i - t_j$. These hyperplanes partition our search space P into polyhedral cells, where the points in each cell lie on the same side of each hyperplane (either on the hyperplane or in one of the two open-halfspaces). It is easy to show that, for every valuation $\mathbf{v} \in V$, all price vectors $\mathbf{p}$ in the same cell yield the same outcome $\mathcal{I}(\mathbf{v},\mathbf{p})$.

[17] Similar geometric partitions have been used previously in various areas. Hartline and Koltun (2005) study item pricing for items with unlimited supply and many unit-demand buyers with known valuations (the model can be viewed also as having one unit-demand buyer with an explicitly given correlated distribution), and give an approximation algorithm, which considers a certain set of price vectors (a suitable grid) and for each price vector partitions the space of valuations by hyperplanes so that valuations in the same cell buy the same item.

Fix a cell C. Since all $\mathbf{p} \in C$ have the same $\mathcal{I}(\mathbf{v}, \mathbf{p})$, we can define $\gamma_i(C)$ as the probability of picking item i with respect to any price vector $\mathbf{p} \in C$. Following the argument used in the proof of Lemma 3.1 we can compute $\gamma_i(C)$ in polynomial time and $\gamma_i(C)$ are rational numbers with polynomial number of bits.

Finally, the supremum of the expected revenue $\mathcal{R}(\mathbf{p})$ over all $\mathbf{p} \in C$ is the maximum of $\sum_i \gamma_i(C) \cdot p_i$ over all $\mathbf{p}$ in the *closure* of C. Let C' denote the closure of C; this is the polyhedron obtained by changing all the strict inequalities of C into weak inequalities. The supremum of $\sum_i \gamma_i(C) \cdot p_i$ over all points $\mathbf{p} \in C$ can be computed in polynomial time by solving the linear program that maximizes $\sum \gamma_i(C) \cdot p_i$ subject to $\mathbf{p} \in C'$. The optimal solution $\mathbf{p}^*_C$ of the LP for cell C is a rational vector whose entries have polynomial number of bits in the number of bits needed to represent the LP (i.e., its number of variables and constraints and the bit-size of the coefficients in the LP), which is polynomial in the size of the input instance.

It follows from an argument similar to the proof of Lemma 2.3 that the revenue $\mathcal{R}(\mathbf{p}^*_C)$ obtained by $\mathbf{p}^*_C$ is at least as large as the supremum of $\mathcal{R}(\mathbf{p})$ over $\mathbf{p} \in C$. As a result, picking the best solution $\mathbf{p}^*_C$ over all cells C (i.e., the one with maximum $\mathcal{R}(\mathbf{p}^*_C)$) gives an optimal price vector $\mathbf{p}^*$ for the given instance. Thus, there is an optimal price vector with polynomial bit complexity. This finishes the proof of the theorem. □

An immediate consequence of the NP membership is that an optimal price vector can be computed in exponential time and polynomial space. A naive brute-force algorithm is to try all rational price vectors in which the number of bits of the prices is bounded by the polynomial implied by the proof of the theorem, and pick the vector that yields the maximum expected revenue. A better algorithm is to generate all the hyperplanes in the proof of the theorem, consider every cell C in the resulting partition of the space, solve the LP for C to compute the optimal solution vector $\mathbf{p}^*_C$, and pick the vector that maximizes the expected revenue $\mathcal{R}(\mathbf{p}^*_C)$ over all cells C. In fact, as we will show below, the LP for a cell has a certain special form, which allows us to formulate it as a shortest path problem and solve it without using Linear Programming. This graph-theoretic approach gives further useful insights into the pricing problem and the structure of the optimal solution, and yields a simpler, more direct method for enumerating a sufficient set of candidate price vectors that is guaranteed to include an optimal solution, without having to solve any Linear Programs, or linear equations, or even shortest path problems.

Next we describe in more detail how to determine whether a set of equations and inequalities defines a nonempty cell, and how to compute the optimal solution over a nonempty cell. The description of a (candidate) cell C consists of equations and inequalities specifying (1) for each item i, the relation of p_i to every value $s_i \in V_i$, and (2) for each pair of items i, j and each pair of values $s_i \in V_i$ and $t_j \in V_j$, the relation of $p_i - p_j$ to $s_i - t_j$. Construct a weighted directed graph $G = (N, E)$ over $n+1$ nodes $N = \{0, 1, \ldots, n\}$ where nodes $1, \ldots, n$ correspond to the n items. For each inequality of the form $p_i < s_i$ or $p_i \le s_i$, include an edge $(0, i)$ with weight s_i, and call the edge strict or weak accordingly as the inequality is strict or weak. In fact, there is a tightest such inequality (i.e., with the smallest value s_i) since the cell is in P, and it suffices to include the edge for this inequality only. Similarly, for each inequality of the form $p_i > s_i$ or $p_i \ge s_i$ (or only for the tightest such inequality, i.e. the one with the largest value s_i) include an edge $(i, 0)$ with weight $-s_i$. For each inequality of the form $p_i - p_j < s_i - t_j$ or $p_i - p_j \le s_i - t_j$ (or only for the tightest such inequality) include a (strict or weak) edge (j, i) with weight $s_i - t_j$. Similarly, for every inequality of the form $p_i - p_j > s_i - t_j$ or $p_i - p_j \ge s_i - t_j$ (or only for the tightest such inequality) include a (strict or weak) edge (i, j) with weight $t_j - s_i$.

We prove the following connections between $G = (N, E)$ and the cell C:

Lemma 3.2. *1. A set of equations and inequalities defines a nonempty cell if and only if the corresponding graph G does not contain a negative weight cycle or a zero weight cycle with a strict edge.*

2. The supremum of the expected revenue for a nonempty cell is bounded from above by the expected revenue $\mathcal{R}(\mathbf{p})$ of the vector $\mathbf{p}$ that consists of the distances from node 0 to the other nodes of the graph G.

Proof. 1. Considering node 0 as having an associated variable p_0 with fixed value 0, the given set of equations (i.e. pairs of weak inequalities) and (strict) inequalities can be viewed as a set of difference constraints on the variables $(p_0, p_1, \ldots, p_n)$, and it is well known that the feasibility of such a set of constraints can be formulated as a negative weight cycle problem. If there is a cycle with negative weight w, then adding all the inequalities corresponding to the edges of the cycle yields the constraint $0 \le w$ (which is false); if there is a cycle with zero weight but also a strict edge, then summing the inequalities yields $0 < 0$.

Conversely, suppose that G does not contain a negative weight cycle or a zero weight cycle with a strict edge. For each strict edge e, replace its weight $w(e)$ by $w'(e) = w(e) - \epsilon$ for a sufficiently small $\epsilon > 0$ (we can treat ϵ symbolically), and let $G(\epsilon)$ be the resulting weighted graph. Note that $G(\epsilon)$ does not contain any negative weight cycle, hence all shortest paths are well-defined in $G(\epsilon)$. Compute the shortest (minimum weight) paths from node 0 to all the other nodes in $G(\epsilon)$, and let $\mathbf{p}(\epsilon)$ be the vector of distances from 0. For each edge (i, j) the distances $p_i(\epsilon)$ and $p_j(\epsilon)$ (where $p_0(\epsilon) = 0$) must satisfy $p_j(\epsilon) \le p_i(\epsilon) + w'(i, j)$, hence all the (weak and strict) inequalities are satisfied.

To determine if a set of equations and inequalities defines a nonempty cell, we can form the graph $G(\epsilon)$ and test for the existence of a negative weight cycle using for example the Bellman–Ford algorithm.

2. Suppose that cell C specified by the constraints is nonempty. Then we claim that the vector $\mathbf{p} = \mathbf{p}(0)$ of distances from node 0 to the other nodes in the graph G is greater than or equal to any vector $\mathbf{p}' \in C$ in all coordinates. We can

show this by induction on the depth of a node in the shortest path tree T of G rooted at node 0. Letting $p'_0 = p_0 = 0$, the basis is trivial. For the induction step, consider a node j with parent i in T. By the inductive hypothesis $p'_i \le p_i$. The edge (i, j) implies that $p'_j - p'_i \le w(i, j)$ or $< w(i, j)$, and the presence of the edge (i, j) in the shortest path tree implies that $p_j = p_i + w(i, j)$. Therefore, $p'_j \le p_j$.

Recall that $\mathcal{R}(\mathbf{p}')$ of $\mathbf{p}' \in C$ is given by $\sum_i \gamma_i(C) \cdot p'_i$. Since $\gamma_i(C)$'s are nonnegative and $\mathbf{p}$ lies in the closure of C, the supremum of $\mathcal{R}(\mathbf{p}')$ over $\mathbf{p}' \in C$ is $\sum_i \gamma_i(C) \cdot p_i$. It follows from the same argument used in the proof of Lemma 2.3 that $\mathcal{R}(\mathbf{p}) \ge \sum_i \gamma_i(C) \cdot p_i$. This finishes the proof of the lemma. $\square$

The NP characterization of Item-Pricing-Decision together with the corresponding structural characterization of Lemma 3.2 for the optimal price vector $\mathbf{p} = \mathbf{p}(0)$ of each cell have several easy and useful consequences.

First, we get an alternative proof of Lemma 2.3 regarding the maximum tie-breaking rule:

Second proof of Lemma 2.3. Suppose that the supremum of the expected revenue is achieved by the supremum within cell C. Let G be the corresponding graph, and let $\mathbf{p}$ be the vector of the distances from node 0 to the other nodes. If $\mathbf{p} \in C$ then the conclusion is immediate, so assume that $\mathbf{p} \notin C$. From the proof of the above lemma we have that $\mathbf{p} \ge \mathbf{p}'$ coordinate-wise for all $\mathbf{p}' \in C$.

We claim that for any valuation $\mathbf{v} \in V$, the revenue $\mathcal{R}(\mathbf{v}, \mathbf{p})$ is at least as large as the revenue $\mathcal{R}(\mathbf{v}, \mathbf{p}')$ under any $\mathbf{p}' \in C$. Suppose that the buyer selects item i under $\mathbf{v}$ for prices $\mathbf{p}'$. Then $p'_i \le v_i$ and thus also $p_i \le v_i$ (since $\mathbf{p}$ is in the closure of C) and thus i is also eligible for selection under $\mathbf{p}$. If the buyer selects i under $\mathbf{p}$ then we know that $p_i \ge p'_i$ and the conclusion follows. Suppose that the buyer selects another item j under $\mathbf{p}$ and that $p'_i > p_j$ and hence $p_i > p_j$. Then we must have $v_j - p_j > v_i - p_i$ due to the tie-breaking rule. The facts that $\mathbf{p}$ is in the closure of C and $v_j - p_j > v_i - p_i$ imply that $v_j - p'_j > v_i - p'_i$ for all $\mathbf{p}' \in C$ and therefore the buyer should have picked j instead of i under prices $\mathbf{p}'$, a contradiction.

We conclude that for any $\mathbf{v} \in V$, $\mathcal{R}(\mathbf{v}, \mathbf{p}) \ge \mathcal{R}(\mathbf{v}, \mathbf{p}')$ for any $\mathbf{p}' \in C$, and the lemma follows. $\square$

Another consequence suggested by the structural characterization of Lemma 3.2 is that the maximum of expected revenue can always be achieved by a price vector $\mathbf{p}$ in which all prices p_i are sums of a value and differences between pairs of values of items. This implies for example the following useful corollary.

Corollary 3.1. *If all the values in V_i, $i \in [n]$, are integers, then there exists an optimal price vector $\mathbf{p} \in P$ with integer coordinates, where $P = \times_{i=1}^{n} [a_i, b_i]$, $a_i = \min_{v \in V_i} v$ and $b_i = \max_{v \in V_i} v$ for each i.*

Furthermore, the following simple algorithm computes an optimal solution by generating a sufficient set of candidate price vectors and picking the best among them. Let $\mathcal{T}$ be the set of all spanning trees of the complete graph on $n+1$ nodes $\{0, 1, \dots, n\}$, where we root every tree at node 0. In the following, $V_j - V_i$ denotes the set $\{v_j - v_i : v_i \in V_i, v_j \in V_j\}$. We let $V_0 = \{0\}$, thus $V_j - V_0 = V_j$.

Algorithm 1

1. Set $r = 0$ and $\mathbf{p}^*$ to be the all-0 vector
2. For every $T \in \mathcal{T}$, for every assignment to each edge (i, j) in T of a weight $w_{ij} \in V_j - V_i$
3.   Compute for each $i \in [n]$ the weight p_i of the path from 0 to i in the weighted tree (T, w)
4.   If $a_i \le p_i \le b_i$ for all $i \in [n]$ then
5.   Compute the expected revenue $\mathcal{R}(\mathbf{p})$ of $\mathbf{p} = (p_1, \dots, p_n)$
6.   If $\mathcal{R}(\mathbf{p}) > r$ then set $r = \mathcal{R}(\mathbf{p})$ and $\mathbf{p}^* = \mathbf{p}$
7. Return $\mathbf{p}^*$ and r

The fact that the set of price vectors generated by the algorithm includes an optimal vector follows from Lemma 3.2. Suppose that C is a cell whose LP yields an optimal vector $\mathbf{p}^*$. Let G be the weighted graph corresponding to C, and let T be the shortest path tree from the node 0 for G. By Lemma 3.2, the optimal price vector for the LP for C consists of the distances from node 0 to the other nodes, which are precisely the weights of the paths in the shortest path tree T from 0 to the other nodes. Algorithm 1 will generate in some iteration this tree T with edge weights as in G, and hence it will generate the optimal price vector $\mathbf{p}^*$.

The algorithm runs in general in exponential time and polynomial space (since the same space is reused for every weighted tree (T, w)). Suppose that the size of the support of every item is at most m. There are $(n+1)^{n-1}$ spanning trees on $n+1$ nodes by Cayley's formula. Each tree has n edges. For the edges of the tree incident to node 0 there are m choices for the weight, and for the other edges there are m^2 choices. Thus, there are less than $(n+1)^{n-1} \cdot m^{2n-1}$ choices of a weighted tree (T, w). For each (T, w), computing the vector $\mathbf{p}$ of the path weights from node 0 to the other nodes takes $O(n)$ time, and it is easy to compute the revenue $\mathcal{R}(\mathbf{p})$ in $O(n^2 m)$ time. Thus, the total running time of the algorithm is $O(n^{n+1} m^{2n})$. Note that if the number of items is constant c, then the time is polynomial, $O(m^{2c})$.

Theorem 3.1. *Algorithm 1 computes an optimal price vector in polynomial space and exponential time. If the number of items is constant, the time is polynomial.*

Remark 1. All the results of this section hold also for correlated (discrete) distributions $\mathcal{F}$, which are given explicitly by listing valuation vectors in the support of $\mathcal{F}$ and their probabilities. In this case, the expected revenue of a given price vector $\mathbf{p}$ can be computed easily by examining every valuation in the support of $\mathcal{F}$, thus Lemma 3.1 holds trivially. It is straightforward to check that all the proofs of other results hold for correlated distributions. In the proof of Theorem 1 the only difference is in the computation of probabilities $\gamma_i(C)$ that item i is selected if the prices are in cell C, which in the case of an explicitly given (discrete) distribution $\mathcal{F}$ is trivial. Thus, Item-Pricing-Decision for correlated distributions $\mathcal{F}$ is in NP. Since the correlated case is known to be NP-hard (Guruswami et al., 2005), the problem is NP-complete. Lemma 3.2, Corollary 3.1, Algorithm 1, and Theorem 3.1 do not depend on $\mathcal{F}$ being a product distribution and apply to arbitrary $\mathcal{F}$.

4. A polynomial-time algorithm for support size 2

In this section, we present a polynomial-time algorithm for the case that each distribution has support size at most 2, thus showing:

Theorem 2. Item-Pricing-Opt *can be solved in polynomial time when all distributions have support size at most* 2*.*

In Section 4.1, we give a polynomial-time algorithm under a certain "non-degeneracy" assumption on the values. In Section 4.2 we use this algorithm to handle the general case, by first perturbing the input instance so that the resulted instance satisfies the "non-degeneracy" assumption, and then solving the latter to obtain an (exactly) optimal price vector of the original instance. Note that we will be able to obtain an exactly optimal solution, instead of an approximate one; this is due to the fact that Section 4.1 not only gives a polynomial-time algorithm for the special case but describes a small, explicit set of $O(n^2)$ price vectors that guarantees to contain all optimal price vectors.

4.1. An interesting special case

In this subsection, we assume that every item has support size 2, where $V_i = \{a_i, b_i\}$ satisfies $b_i > a_i > 0$, for all $i \in [n]$. Let $q_i : 0 < q_i < 1$ denote the probability of the value of item i being b_i. For convenience, we also let $t_i = b_i - a_i > 0$. In addition, we assume in this subsection that the value-vectors $\mathbf{a} = (a_1, \ldots, a_n)$ and $\mathbf{b} = (b_1, \ldots, b_n)$ satisfy the following "non-degeneracy" assumption:

Non-degeneracy assumption. $b_1 < b_2 < \cdots < b_n$, $a_i \neq a_j$ and $t_i \neq t_j$ for all $i, j \in [n]$.

As we show next in Section 4.2, this special case encapsulates the essential difficulty of the problem.

Let OPT denote the set of optimal price vectors in $P = \times_{i=1}^{n}[a_i, b_i]$ that maximize the expected revenue $\mathcal{R}(\mathbf{p})$. (By Lemma 2.2 the set OPT is nonempty.) Next we prove a sequence of lemmas to impose more and more stringent conditions on the set of optimal price vectors. The proof of each lemma in the sequence shares a similar flavor, where we show that if a price vector violates the condition then one can modify it to achieve a strictly higher revenue. At the end we establish that, given $\mathbf{a}$ and $\mathbf{b}$ that satisfy all the conditions above, one can compute efficiently a set $A \subseteq P$ of price vectors such that $|A| = O(n^2)$ and $\mathrm{OPT} \subseteq A$ (see below for an explicit description of A). By computing $\mathcal{R}(\mathbf{p})$ for every $\mathbf{p} \in A$ (using Lemma 3.1), we get both the maximum of expected revenue and an optimal price vector.

We outline the main steps of the proof. First we show that an optimal price vector $\mathbf{p} \in P$ is either equal to $\mathbf{b}$ or has at least one coordinate p_k equal to the low value a_k. Second we show that in the latter case there must be exactly one k such that $p_k = a_k$. Third, we show that all other coordinates p_i must be either equal to b_i or to $b_i - t_k$ where $t_k = b_k - a_k$. Furthermore, (1) for all $i < k$, we must have $p_i = b_i$; (2) if $i > k$ and $t_i < t_k$ then $b_i - t_k < a_i$, and hence p_i must be equal to b_i. Thus, the question of whether p_i should be b_i or $b_i - t_k$ concerns only the subset T_k of items $i > k$ such that $t_i > t_k$ (note that $t_i \neq t_k$ by the non-degeneracy assumption). The final lemma, which is also the most difficult one in the sequence, shows that the choice of an optimal vector $\mathbf{p}$ must be monotonic: if it sets $p_i = b_i - t_k$ for some i, then it must set $p_j = b_j - t_k$ for all $j > i$, $j \in T_k$. As a consequence, there are only $O(n^2)$ price vectors that can possibly be optimal: (1) $\mathbf{b}$, (2) for each $k \in [n]$, the vector that has $p_k = a_k$ and $p_i = b_i$ for all $i \neq k$, and (3) for each $k \in [n]$ and each $i > k, i \in T_k$, the vector that has $p_k = a_k$, $p_j = b_j - t_j$ for all $j > i$, $j \in T_k$, and $p_j = b_j$ for all other j.

We proceed now with the detailed proof. We start with the following lemma:

Lemma 4.1. *If* $\mathbf{p} \in P$ *satisfies* $p_i > a_i$ *for all* $i \in [n]$*, then either* $\mathbf{p} = \mathbf{b}$ *or we have* $\mathbf{p} \notin \mathrm{OPT}$.

Proof. Assume for contradiction that $\mathbf{p} \in P$ satisfies $p_i > a_i$, for all $i \in [n]$ but $\mathbf{p} \neq \mathbf{b}$. It then follows from the maximum price tie-breaking rule that $\mathcal{R}(\mathbf{v}, \mathbf{b}) \geq \mathcal{R}(\mathbf{v}, \mathbf{p})$ for all $\mathbf{v} \in V$. Moreover, there is at least one $\mathbf{v}^* \in V$ such that $\mathcal{R}(\mathbf{v}^*, \mathbf{b}) > \mathcal{R}(\mathbf{v}^*, \mathbf{p})$:

If $p_i < b_i$, then consider $\mathbf{v}^*$ with $v_i^* = b_i$ and $v_j^* = a_j$ for all other j. It follows that $\mathcal{R}(\mathbf{b}) > \mathcal{R}(\mathbf{p})$ as we assumed that $0 < q_i < 1$ for all $i \in [n]$ and thus, $\mathbf{p} \notin \mathsf{OPT}$. □

Next we show that there can be at most one i such that $p_i = a_i$; otherwise $\mathbf{p} \notin \mathsf{OPT}$. We emphasize that all the conditions on V_i are assumed in the lemmas below, the non-degeneracy assumption in particular.

Lemma 4.2. *If $\mathbf{p} \in P$ has more than one $i \in [n]$ such that $p_i = a_i$, then we have $\mathbf{p} \notin \mathsf{OPT}$.*

Proof. Assume for contradiction that $\mathbf{p} \in P$ has more than one i such that $p_i = a_i$. We prove the lemma by explicitly constructing a new price vector $\mathbf{p}' \in P$ from $\mathbf{p}$ such that $\mathcal{R}(\mathbf{v}, \mathbf{p}') \geq \mathcal{R}(\mathbf{v}, \mathbf{p})$ for all $\mathbf{v} \in V$ and $\mathcal{R}(\mathbf{v}^*, \mathbf{p}') > \mathcal{R}(\mathbf{v}^*, \mathbf{p})$ for at least one $\mathbf{v}^* \in V$. This implies that $\mathcal{R}(\mathbf{p}') > \mathcal{R}(\mathbf{p})$ and thus, $\mathbf{p}$ is not optimal. We will be using this simple strategy in most of the proofs of this section.

Let $k \in [n]$ denote *the* item with the smallest a_k among all $i \in [n]$ with $p_i = a_i$. By the non-degeneracy assumption, k is unique. Recall that $t_k = b_k - a_k = b_k - p_k$. We let S denote the set of $i \in [n]$ such that $b_i - p_i = t_k$, so $k \in S$. By the non-degeneracy assumption again, we have $p_i > a_i$ for all $i \in S - \{k\}$. We now construct $\mathbf{p}' \in P$ as follows: For each $i \in [n]$, set $p_i' = p_i$ if $i \notin S$; otherwise set $p_i' = p_i + \epsilon$ for some sufficiently small $\epsilon > 0$. Next we show that $\mathcal{R}(\mathbf{v}, \mathbf{p}') \geq \mathcal{R}(\mathbf{v}, \mathbf{p})$ for all $\mathbf{v} \in V$. Fix a $\mathbf{v} \in V$. We consider the following three cases:

1. If $\mathcal{U}(\mathbf{v}, \mathbf{p}) = t_k$, then $\mathcal{T}(\mathbf{v}, \mathbf{p}) \subseteq S$ by the definition of S. When ϵ is sufficiently small, we have
$$\mathcal{T}(\mathbf{v}, \mathbf{p}') = \mathcal{T}(\mathbf{v}, \mathbf{p}) \quad \text{and} \quad \mathcal{R}(\mathbf{v}, \mathbf{p}') = \mathcal{R}(\mathbf{v}, \mathbf{p}) + \epsilon > \mathcal{R}(\mathbf{v}, \mathbf{p}).$$
2. If $\mathcal{U}(\mathbf{v}, \mathbf{p}) = 0$ and $k \in \mathcal{T}(\mathbf{v}, \mathbf{p})$, then we have $\mathcal{T}(\mathbf{v}, \mathbf{p}) \cap S = \{k\}$ since $b_i > p_i > a_i$ for all other $i \in S$. We claim that $\mathcal{R}(\mathbf{v}, \mathbf{p}) > p_k$ in this case. To see this, note that there exists an item $\ell \in [n]$ such that $p_\ell = a_\ell$ and $p_\ell > p_k$ by our choice of k. As $\mathcal{U}(\mathbf{v}, \mathbf{p}) = 0$, we must have $v_\ell = a_\ell$ and thus, $\ell \in \mathcal{T}(\mathbf{v}, \mathbf{p})$ and $\mathcal{R}(\mathbf{v}, \mathbf{p}) \geq p_\ell$ is not obtained from selling item k. Therefore, we have
$$\mathcal{U}(\mathbf{v}, \mathbf{p}') = 0, \quad \mathcal{T}(\mathbf{v}, \mathbf{p}') = \mathcal{T}(\mathbf{v}, \mathbf{p}) - \{k\} \quad \text{and} \quad \mathcal{R}(\mathbf{v}, \mathbf{p}') = \mathcal{R}(\mathbf{v}, \mathbf{p}).$$
3. Finally, if neither of the cases above happens, then we have $\mathcal{T}(\mathbf{v}, \mathbf{p}) \cap S = \emptyset$ (note that this includes the case when $\mathcal{T}(\mathbf{v}, \mathbf{p}) = \emptyset$). For this case we have $\mathcal{T}(\mathbf{v}, \mathbf{p}') = \mathcal{T}(\mathbf{v}, \mathbf{p})$ and $\mathcal{R}(\mathbf{v}, \mathbf{p}') = \mathcal{R}(\mathbf{v}, \mathbf{p})$.

The lemma then follows because in the second case above, we indeed showed that the following valuation vector $\mathbf{v}^*$ in V satisfies $\mathcal{R}(\mathbf{v}^*, \mathbf{p}') > \mathcal{R}(\mathbf{v}^*, \mathbf{p})$: $v_k = b_k$ and $v_i = a_i$ for all $i \neq k$. □

Lemma 4.2 reduces our search space to $\mathbf{p}$ such that either $\mathbf{p} = \mathbf{b}$ or $\mathbf{p} \in P_k$ for some $k \in [n]$, where we use P_k to denote the set of price vectors $\mathbf{p} \in P$ such that $p_k = a_k$ and $p_i > a_i$ for all other $i \in [n]$.

The next lemma further restricts our attention to $\mathbf{p} \in P_k$ such that $p_i \in \{b_i, b_i - t_k\}$ for all $i \neq k$.

Lemma 4.3. *If $\mathbf{p} \in P_k$ but $p_i \notin \{b_i, b_i - t_k\}$ for some $i \neq k$, then we have $\mathbf{p} \notin \mathsf{OPT}$.*

Proof. Assume for contradiction that $p_\ell \notin \{b_\ell, b_\ell - t_k\}$. As $\mathbf{p} \in P_k$, we also have $p_\ell > a_\ell$. Now we use S to denote the set of all $i \in [n]$ such that $b_i - p_i = b_\ell - p_\ell$. It is clear that $k \notin S$. We use $\mathbf{p}'$ to denote the following new price vector: $p_i' = p_i$ for all $i \notin S$, and $p_i' = p_i + \epsilon$ for all $i \in S$, where $\epsilon > 0$ is sufficiently small. We use the same proof strategy to show that $\mathcal{R}(\mathbf{p}') > \mathcal{R}(\mathbf{p})$. For each valuation $\mathbf{v} \in V$, we have

1. If $\mathcal{U}(\mathbf{v}, \mathbf{p}) < 0$, then clearly $\mathcal{U}(\mathbf{v}, \mathbf{p}') < 0$ as well and thus, $\mathcal{R}(\mathbf{v}, \mathbf{p}') = \mathcal{R}(\mathbf{v}, \mathbf{p}) = 0$.
2. If $\mathcal{U}(\mathbf{v}, \mathbf{p}) = b_\ell - p_\ell$, then $\mathcal{T}(\mathbf{v}, \mathbf{p}) \subseteq S$ by the definition of S. When ϵ is sufficiently small,
$$\mathcal{T}(\mathbf{v}, \mathbf{p}') = \mathcal{T}(\mathbf{v}, \mathbf{p}) \quad \text{and} \quad \mathcal{R}(\mathbf{v}, \mathbf{p}') = \mathcal{R}(\mathbf{v}, \mathbf{p}) + \epsilon > \mathcal{R}(\mathbf{v}, \mathbf{p}).$$
3. If $\mathcal{U}(\mathbf{v}, \mathbf{p}) \geq 0$ but $\mathcal{U}(\mathbf{v}, \mathbf{p}) \neq b_\ell - p_\ell$, then it is easy to see that $\mathcal{T}(\mathbf{v}, \mathbf{p}) \cap S = \emptyset$, because $p_i > a_i$ and $b_i - p_i = b_\ell - p_\ell$ for all $i \in S$. It follows that $\mathcal{T}(\mathbf{v}, \mathbf{p}') = \mathcal{T}(\mathbf{v}, \mathbf{p})$ and $\mathcal{R}(\mathbf{v}, \mathbf{p}') = \mathcal{R}(\mathbf{v}, \mathbf{p})$.

The lemma follows by combining all three cases. □

As suggested by Lemma 4.3, for each $k \in [n]$, we use P_k' to denote the set of $\mathbf{p} \in P_k$ such that $p_k = a_k$ and $p_i \in \{b_i, b_i - t_k\}$ for all other i. In particular, p_i must be b_i if $t_i < t_k$ ($t_i \neq t_k$, by the non-degeneracy assumption). The next lemma shows that we only need to consider $\mathbf{p} \in P_k'$ such that $p_i = b_i$ for all $i < k$.

Lemma 4.4. *If $\mathbf{p} \in P_k'$ satisfies $p_\ell = b_\ell - t_k > a_\ell$ for some $\ell < k$, then we have $\mathbf{p} \notin \mathsf{OPT}$.*

Proof. We construct $\mathbf{p}'$ from $\mathbf{p}$ as follows. Let S denote the set of all $i < k$ such that $p_i = b_i - t_k > a_i$. By our assumption, S is nonempty. Then set $p'_i = p_i$ for all $i \notin S$ and $p'_i = p_i + \epsilon$ for all $i \in S$, where $\epsilon > 0$ is sufficiently small. Similarly we show that $\mathcal{R}(\mathbf{p}') > \mathcal{R}(\mathbf{p})$ by considering the following subcases:

1. If $\mathcal{U}(\mathbf{v}, \mathbf{p}) = t_k$ and $\mathcal{T}(\mathbf{v}, \mathbf{p}) \cap S \neq \emptyset$, we consider the following cases. If $\mathcal{T}(\mathbf{v}, \mathbf{p}) \subseteq S$, then
$$\mathcal{T}(\mathbf{v}, \mathbf{p}') = \mathcal{T}(\mathbf{v}, \mathbf{p}) \quad \text{and} \quad \mathcal{R}(\mathbf{v}, \mathbf{p}') = \mathcal{R}(\mathbf{v}, \mathbf{p}) + \epsilon > \mathcal{R}(\mathbf{v}, \mathbf{p}).$$
(Note that there exists a valuation $\mathbf{v}$ that falls in this subcase, for example, the valuation $\mathbf{v}$ that has $v_i = b_i$ for $i \in S$ and $v_i = a_i$ for $i \notin S$.) Otherwise, there exists a $j \geq k$ such that $j \in \mathcal{T}(\mathbf{v}, \mathbf{p})$. This implies that $\mathcal{R}(\mathbf{v}, \mathbf{p}) \geq p_j = b_j - t_k$ is not obtained from any item in S. As a result, we have
$$\mathcal{T}(\mathbf{v}, \mathbf{p}') = \mathcal{T}(\mathbf{v}, \mathbf{p}) - S \quad \text{and} \quad \mathcal{R}(\mathbf{v}, \mathbf{p}') = \mathcal{R}(\mathbf{v}, \mathbf{p}).$$
2. If the case above does not happen, then $\mathcal{T}(\mathbf{v}, \mathbf{p}) \cap S = \emptyset$. As a result, we have
$$\mathcal{T}(\mathbf{v}, \mathbf{p}') = \mathcal{T}(\mathbf{v}, \mathbf{p}) \quad \text{and} \quad \mathcal{R}(\mathbf{v}, \mathbf{p}') = \mathcal{R}(\mathbf{v}, \mathbf{p}).$$

The lemma follows by combining the two cases. □

Finally, we use P_k^* for each $k \in [n]$ to denote the set of $\mathbf{p} \in P$ such that $p_k = a_k$; $p_i = b_i$ for all $i < k$; $p_i = b_i$, for all $i > k$ such that $t_i < t_k$; and $p_i \in \{b_i, b_i - t_k\}$, for all other $i > k$. However, P_k^* may still be exponentially large in general. Let T_k denote the set of $i > k$ such that $t_i > t_k$. Given $\mathbf{p} \in P_k^*$, our last lemma below implies that, if i is the smallest index in T_k such that $p_i = b_i - t_k$, then $p_j = b_j - t_k$ for all $j \in T_k$ larger than i; otherwise $\mathbf{p}$ is not optimal. In other words, $\mathbf{p}$ has to be monotone in setting p_j, $j \in T_k$, to be $b_j - t_k$; otherwise $\mathbf{p}$ is not optimal. As a result, there are only $O(n^2)$ many price vectors that we need to check, and the best one among them is optimal. We use $A \subseteq \cup_k P_k^*$ to denote this set of price vectors.

Lemma 4.5. *Given $k \in [n]$ and $\mathbf{p} \in P_k^*$, if there exist two indices $c, d \in T_k$ such that $c < d$, $p_c = b_c - t_k$ but $p_d = b_d$, then we must have $\mathbf{p} \notin \mathsf{OPT}$.*

Proof. We use t to denote t_k for convenience. Also we may assume, without loss of generality, that there is no index between c and d in T_k; otherwise we can use it to replace either c or d, depending on its price.

We define two vectors from $\mathbf{p}$. First, let $\mathbf{p}'$ denote the vector obtained from $\mathbf{p}$ by replacing $p_d = b_d$ by $p'_d = b_d - t$. Let $\mathbf{p}^*$ denote the vector obtained from $\mathbf{p}$ by replacing $p_c = b_c - t$ by $p^*_c = b_c$. In other words, the cth and dth entries of $\mathbf{p}, \mathbf{p}', \mathbf{p}^*$ are $(b_c - t, b_d)$, $(b_c - t, b_d - t)$, (b_c, b_d), respectively, while all other $n-2$ entries are the same. Our plan is to show that if $\mathcal{R}(\mathbf{p}) \geq \mathcal{R}(\mathbf{p}')$, then $\mathcal{R}(\mathbf{p}^*) > \mathcal{R}(\mathbf{p})$. This implies that $\mathbf{p}$ cannot be optimal and the lemma follows.

We need some notation. Let V' denote the projection of V onto all but the cth and dth coordinates:
$$V' = \times_{i \in [n] - \{c,d\}} V_i.$$
We use $[n] - \{c, d\}$ to index entries of vectors $\mathbf{u}$ in V'. Let $U \subseteq V'$ denote the set of vectors $\mathbf{u} \in V'$ such that $u_i - p_i < t$ for all $i > d$. (This just means that for each $i \in T_k$, if $i > d$ and $p_i = b_i - t$, then $u_i = a_i$.) Given $\mathbf{u} \in V'$, $v_c \in \{a_c, b_c\}$ and $v_d \in \{a_d, b_d\}$, we use $(\mathbf{u}, v_c, v_d)$ to denote a n-dimensional price vector in V. Now we compare the expected revenue $\mathcal{R}(\mathbf{p})$, $\mathcal{R}(\mathbf{p}')$ and $\mathcal{R}(\mathbf{p}^*)$.

First, we claim that, if $\mathbf{v} = (\mathbf{u}, v_c, v_d) \in V$ but $\mathbf{u} \notin U$, then we have $\mathcal{R}(\mathbf{v}, \mathbf{p}) = \mathcal{R}(\mathbf{v}, \mathbf{p}') = \mathcal{R}(\mathbf{v}, \mathbf{p}^*)$. This is simply because there exists an item $i > d$ such that $v_i - p_i = t$, so it always dominates both items c and d. As a result, the difference among $\mathbf{p}, \mathbf{p}'$ and $\mathbf{p}^*$ no longer matters. Second, it is easy to show that for any $\mathbf{v} = (\mathbf{u}, a_c, a_d) \in V$, then $\mathcal{R}(\mathbf{v}, \mathbf{p}) = \mathcal{R}(\mathbf{v}, \mathbf{p}') = \mathcal{R}(\mathbf{v}, \mathbf{p}^*)$ as the utility from c and d are negative.

Now we consider a vector $\mathbf{v} = (\mathbf{u}, v_c, v_d) \in V$ such that $\mathbf{u} \in U$ and (v_c, v_d) is either (a_c, b_d), (b_c, a_d), or (b_c, b_d). For convenience, for each $\mathbf{u} \in U$ we use $\mathbf{u}_1^+$ to denote $(\mathbf{u}, a_c, b_d)$; $\mathbf{u}_2^+$ to denote $(\mathbf{u}, b_c, a_d)$; and $\mathbf{u}_3^+$ to denote $(\mathbf{u}, b_c, b_d)$. By the definition of U, we have the following simple cases:

1. For $\mathbf{p}$, we have $\mathcal{R}(\mathbf{u}_2^+, \mathbf{p}) = b_c - t$ and $\mathcal{R}(\mathbf{u}_3^+, \mathbf{p}) = b_c - t$;
2. For $\mathbf{p}'$, we have $\mathcal{R}(\mathbf{u}_1^+, \mathbf{p}') = b_d - t$, $\mathcal{R}(\mathbf{u}_2^+, \mathbf{p}') = b_c - t$ and $\mathcal{R}(\mathbf{u}_3^+, \mathbf{p}') = b_d - t$.

We need the following equation:
$$\mathcal{R}(\mathbf{u}_1^+, \mathbf{p}) = \mathcal{R}(\mathbf{u}_1^+, \mathbf{p}^*) = \mathcal{R}(\mathbf{u}_3^+, \mathbf{p}^*). \tag{2}$$
It holds because in all three cases, the buyer prefers item d, among items c and d, with utility 0 and price b_d. Similarly we also have the following two inequalities:

$$\mathcal{R}(\mathbf{u}_1^+, \mathbf{p}^*) - (b_d - b_c) \le \mathcal{R}(\mathbf{u}_2^+, \mathbf{p}^*) \le \mathcal{R}(\mathbf{u}_1^+, \mathbf{p}^*). \tag{3}$$

The second inequality holds because in $(\mathbf{u}_2^+, \mathbf{p}^*)$, the buyer prefers item c, among c and d, with utility 0 and price b_c, while in $(\mathbf{u}_1^+, \mathbf{p}^*)$, the buyer prefers item d with utility 0 and price b_d. On the other hand, the first inequality holds because, either (1) there is an item with a positive utility or utility 0 and price $> p_d$, in which case the buyer will not pick items c or d in both situations and we have $\mathcal{R}(\mathbf{u}_2^+, \mathbf{p}^*) = \mathcal{R}(\mathbf{u}_1^+, \mathbf{p}^*)$, or (2) there is no such item and thus, the buyer selects item d in $(\mathbf{u}_1^+, \mathbf{p}^*)$ with $\mathcal{R}(\mathbf{u}_1^+, \mathbf{p}^*) = b_d$ and at the same time, we have $\mathcal{R}(\mathbf{u}_2^+, \mathbf{p}^*) \ge b_c$.

Given a $\mathbf{v} \in V$, recall that $\Pr[\mathbf{v}]$ denotes the probability of the valuation vector being $\mathbf{v}$. Given a $\mathbf{u} \in U$, we also use $\Pr[\mathbf{u}]$ to denote the probability of the $n-2$ items, except items c and d, taking values $\mathbf{u}$. Let

$$h_1 = (1-q_c)q_d, \quad h_2 = q_c(1-q_d) \quad \text{and} \quad h_3 = q_c q_d.$$

Clearly we have $h_1, h_2, h_3 > 0$ and $\Pr[\mathbf{u}_i^+] = \Pr[\mathbf{u}] \cdot h_i$, for all $\mathbf{u} \in U$ and $i \in [3]$.

In order to compare $\mathcal{R}(\mathbf{p})$, $\mathcal{R}(\mathbf{p}')$ and $\mathcal{R}(\mathbf{p}^*)$, we only need to compare the following three sums:

$$\sum_{i\in[3]}\sum_{\mathbf{u}\in U} \Pr[\mathbf{u}_i^+] \cdot \mathcal{R}(\mathbf{u}_i^+, \mathbf{p}), \quad \sum_{i\in[3]}\sum_{\mathbf{u}\in U} \Pr[\mathbf{u}_i^+] \cdot \mathcal{R}(\mathbf{u}_i^+, \mathbf{p}') \quad \text{and} \quad \sum_{i\in[3]}\sum_{\mathbf{u}\in U} \Pr[\mathbf{u}_i^+] \cdot \mathcal{R}(\mathbf{u}_i^+, \mathbf{p}^*).$$

For the first sum, we can rewrite it as (here all sums are over $\mathbf{u} \in U$):

$$h_1 \cdot \sum_{\mathbf{u}} \Pr[\mathbf{u}] \cdot \mathcal{R}(\mathbf{u}_1^+, \mathbf{p}) + h_2 \cdot \sum_{\mathbf{u}} \Pr[\mathbf{u}] \cdot (b_c - t) + h_3 \cdot \sum_{\mathbf{u}} \Pr[\mathbf{u}] \cdot (b_c - t), \tag{4}$$

while the sum for $\mathcal{R}(\mathbf{p}')$ is the following:

$$h_1 \cdot \sum_{\mathbf{u}} \Pr[\mathbf{u}] \cdot (b_d - t) + h_2 \cdot \sum_{\mathbf{u}} \Pr[\mathbf{u}] \cdot (b_c - t) + h_3 \cdot \sum_{\mathbf{u}} \Pr[\mathbf{u}] \cdot (b_d - t). \tag{5}$$

Since $c < d$ and $b_c < b_d$, $\mathcal{R}(\mathbf{p}) \ge \mathcal{R}(\mathbf{p}')$ would imply that

$$\sum_{\mathbf{u}} \Pr[\mathbf{u}] \cdot \mathcal{R}(\mathbf{u}_1^+, \mathbf{p}) > \sum_{\mathbf{u}} \Pr[\mathbf{u}] \cdot (b_d - t). \tag{6}$$

On the other hand, we can also rewrite the sum for $\mathcal{R}(\mathbf{p}^*)$ as

$$h_1 \cdot \sum_{\mathbf{u}} \Pr[\mathbf{u}] \cdot \mathcal{R}(\mathbf{u}_1^+, \mathbf{p}^*) + h_2 \cdot \sum_{\mathbf{u}} \Pr[\mathbf{u}] \cdot \mathcal{R}(\mathbf{u}_2^+, \mathbf{p}^*) + h_3 \cdot \sum_{\mathbf{u}} \Pr[\mathbf{u}] \cdot \mathcal{R}(\mathbf{u}_3^+, \mathbf{p}^*). \tag{7}$$

The first sum in (7) is the same as that of (4). For the second sum, from (3), (2) and (6) we have

$$\begin{aligned}\sum_{\mathbf{u}} \Pr[\mathbf{u}] \cdot \mathcal{R}(\mathbf{u}_2^+, \mathbf{p}^*) &\ge \sum_{\mathbf{u}} \Pr[\mathbf{u}] \cdot \Big(\mathcal{R}(\mathbf{u}_1^+, \mathbf{p}) - (b_d - b_c)\Big) \\ &> \sum_{\mathbf{u}} \Pr[\mathbf{u}] \cdot \Big(b_d - t - (b_d - b_c)\Big) = \sum_{\mathbf{u}} \Pr[\mathbf{u}] \cdot (b_c - t).\end{aligned}$$

The third sum in (7) is also strictly larger than that of (4) as $\mathcal{R}(\mathbf{u}_3^+, \mathbf{p}^*) = \mathcal{R}(\mathbf{u}_1^+, \mathbf{p}^*) \ge \mathcal{R}(\mathbf{u}_2^+, \mathbf{p}^*)$ while the second and third sums in (4) are the same, ignoring h_2 and h_3. Thus, $\mathcal{R}(\mathbf{p}^*) > \mathcal{R}(\mathbf{p})$. □

Remark 2. The first four lemmas apply to any distribution $\mathcal{F}$ that is not necessarily a product distribution, as long as it has support $\times_{i=1}^n \{a_i, b_i\}$. Lemma 4.5, however, requires $\mathcal{F}$ to be a product distribution.

4.2. General case

Now we deal with the general case. Let I denote an input instance with n items, in which $|V_i| \le 2$ for all i. For each $i \in [n]$, either $V_i = \{a_i, b_i\}$ where $b_i > a_i \ge 0$, or $V_i = \{b_i\}$, where $b_i \ge 0$. We let $D \subseteq [n]$ denote the set of $i \in [n]$ such that $|V_i| = 2$. For each item $i \in D$, we use $q_i : 0 < q_i < 1$ to denote the probability of its value being b_i. Each item $i \notin D$ has value b_i with probability 1. As permuting the items does not affect the maximum expected revenue, we may assume without loss of generality that $b_1 \le b_2 \le \cdots \le b_n$.

The idea is to perturb I (symbolically), so that the new instances satisfy all conditions described at the beginning of the section, which we know how to solve efficiently. For this purpose, we define a new n-item instance I_ϵ from I for any $\epsilon > 0$: For each $i \in D$, the support of item i is $V_{i,\epsilon} = \{a_i + i\epsilon, b_i + 2i\epsilon\}$, and for each $i \notin D$, the support of item i is $V_{i,\epsilon} = \{b_i + i\epsilon, b_i + 2i\epsilon\}$. For each $i \in D$, the probability of the value being $b_i + 2i\epsilon$ is still set to be q_i, while for each $i \notin D$, the probability of the value being $b_i + 2i\epsilon$ is set to be $1/2$. In the rest of the section, we use $\mathcal{R}(\mathbf{p})$ and $\mathcal{R}(\mathbf{v}, \mathbf{p})$ to denote the revenue with respect to I, and use $\mathcal{R}_\epsilon(\mathbf{p})$ and $\mathcal{R}_\epsilon(\mathbf{v}, \mathbf{p})$ to denote the revenue with respect to I_ϵ. Let $V_\epsilon = \times_{i=1}^n V_{i,\epsilon}$. Let

ρ denote the following map from V_ϵ to V: ρ maps $\mathbf{u} \in V_\epsilon$ to $\mathbf{v} \in V$, where 1) $v_i = b_i$ when $i \notin D$; 2) $v_i = a_i$ if $u_i = a_i + i\epsilon$ and $v_i = b_i$ if $u_i = b_i + 2i\epsilon$ when $i \in D$.

It is easy to verify that, when $\epsilon > 0$ is sufficiently small, the new instance I_ϵ satisfies all conditions given at the beginning of the section, including the non-degeneracy assumption. From the previous subsection we know then that, there is a set of $O(n^2)$ price vectors for I_ϵ, denote this set by A_ϵ, such that the best vector in A_ϵ is optimal for I_ϵ and achieves $\max_{\mathbf{p}} \mathcal{R}_\epsilon(\mathbf{p})$. Furthermore, from the construction of A_ϵ, we know that every vector $\mathbf{p}_\epsilon$ in A_ϵ has an explicit expression in ϵ: each entry of $\mathbf{p}_\epsilon$ is indeed an affine linear function of ϵ. Moreover, we show that

Lemma 4.6. *The limit of* $\max_{\mathbf{p}} \mathcal{R}_\epsilon(\mathbf{p})$ *exists as* $\epsilon \to 0$, *it is equal to*

$$\max_{\mathbf{p}_\epsilon \in A_\epsilon} \left\{ \lim_{\epsilon \to 0} \mathcal{R}_\epsilon(\mathbf{p}_\epsilon) \right\},$$

and it can be computed in polynomial time.

Proof. Since I_ϵ satisfies all the conditions for small enough ϵ, $\max_{\mathbf{p}} \mathcal{R}_\epsilon(\mathbf{p})$ is achieved by one of the $O(n^2)$ price vectors in A_ϵ. The entries of every vector $\mathbf{p}_\epsilon \in A_\epsilon$ are affine linear functions of ϵ. As a result, the limit of $\mathcal{R}_\epsilon(\mathbf{p}_\epsilon)$ as ϵ approaches 0 exists and can be computed efficiently. Since $\lim_{\epsilon \to 0} \big(\max_{\mathbf{p}} \mathcal{R}_\epsilon(\mathbf{p})\big)$ is just the maximum of these $O(n^2)$ limits $\lim_{\epsilon \to 0} \mathcal{R}_\epsilon(\mathbf{p}_\epsilon)$, $\mathbf{p}_\epsilon \in A_\epsilon$, it also exists and can be computed in polynomial time in the input size of I. □

Finally, the next two lemmas show that this limit is exactly the maximum expected revenue of I. This finishes the proof of Theorem 2.

Lemma 4.7. $\max_{\mathbf{p}} \mathcal{R}(\mathbf{p}) \le \lim_{\epsilon \to 0} \big(\max_{\mathbf{p}} \mathcal{R}_\epsilon(\mathbf{p})\big)$.

Proof. Let $\mathbf{p}^*$ denote an optimal price vector of I. It suffices to show that, when ϵ is sufficiently small,

$$\max_{\mathbf{p}} \mathcal{R}_\epsilon(\mathbf{p}) \ge \mathcal{R}(\mathbf{p}^*) - 4n^2\epsilon. \tag{8}$$

The proof is similar to that of Lemma 2.1. Let $\mathbf{p}'$ denote the vector in which $p'_i = \max\big(0, p^*_i - 4r_i n\epsilon\big)$, where r_i is the rank of p^*_i among $\{p^*_1, \ldots, p^*_n\}$ sorted in the increasing order (when there are ties, items with lower index are ranked higher). We claim that, when $\epsilon > 0$ is sufficiently small,

$$\mathcal{R}_\epsilon(\mathbf{u}, \mathbf{p}') \ge \mathcal{R}(\rho(\mathbf{u}), \mathbf{p}^*) - 4n^2\epsilon, \quad \text{for any } \mathbf{u} \in V_\epsilon, \tag{9}$$

from which we get $\mathcal{R}_\epsilon(\mathbf{p}') \ge \mathcal{R}(\mathbf{p}^*) - 4n^2\epsilon$ and (8) follows.

To prove (9), we fix a $\mathbf{u} \in V_\epsilon$ and let $\mathbf{v} = \rho(\mathbf{u}) \in V$. (9) holds trivially if $\mathcal{R}(\mathbf{v}, \mathbf{p}^*) = 0$. Assume that $\mathcal{R}(\mathbf{v}, \mathbf{p}^*) > 0$, and let k denote the item selected in I on $(\mathbf{v}, \mathbf{p}^*)$. (9) also holds trivially if $p^*_k < 4n^2\epsilon$, so without loss of generality, we assume that $p_k \ge 4n^2\epsilon$. For any other item $j \in [n]$, we compare the utilities of items k and j in I_ϵ on $(\mathbf{u}, \mathbf{p}')$. We claim that

$$u_k - p'_k > u_j - p'_j \tag{10}$$

because 1) if $v_k - p^*_k > v_j - p^*_j$, then (10) holds when ϵ is sufficiently small; 2) if $v_k - p^*_k = v_j - p^*_j$ and $p^*_k > p^*_j$, then (10) holds because $p^*_k - p'_k - (p^*_j - p'_j) \ge 4n\epsilon > (v_k - u_k) + (u_j - v_j)$; 3) finally, the case when $v_k - p^*_k = v_j - p^*_j$, $p_k = p_j$ and $k < j$ follows similarly from $r_k > r_j$. Therefore, k remains to be the item being selected in I_ϵ on $(\mathbf{u}, \mathbf{p}')$. (9) then follows from the fact that $p'_k \ge p^*_k - 4n^2\epsilon$ by definition. □

Lemma 4.8. $\max_{\mathbf{p}} \mathcal{R}(\mathbf{p}) \ge \lim_{\epsilon \to 0} \big(\max_{\mathbf{p}} \mathcal{R}_\epsilon(\mathbf{p})\big)$.

Proof. From Lemma 4.6, there is a price vector $\mathbf{p}_\epsilon \in A_\epsilon$ in which every entry is an affine linear function of ϵ, such that (as the cardinality of $|A_\epsilon|$ is bounded from above by $O(n^2)$)

$$\lim_{\epsilon \to 0} \left(\max_{\mathbf{p}} \mathcal{R}_\epsilon(\mathbf{p}) \right) = \lim_{\epsilon \to 0} \mathcal{R}_\epsilon(\mathbf{p}_\epsilon).$$

Let $\widetilde{\mathbf{p}} \in \mathbb{R}^n_+$ denote the limit of $\mathbf{p}_\epsilon$, by simply removing all the ϵ's in the affine linear functions. Moreover, we note that $|\widetilde{p}_i - p_{\epsilon,i}| = O(n\epsilon)$ by the construction of A_ϵ, where we use $p_{\epsilon,i}$ to denote the ith entry of $\mathbf{p}_\epsilon$.

Next, let $\mathbf{q}_\epsilon$ denote the vector in which the ith entry $q_{\epsilon,i} = \max(0, \widetilde{p}_i - r_i n^2\epsilon)$ for all $i \in [n]$, where r_i is the rank of $\widetilde{p}_i$ among entries of $\widetilde{\mathbf{p}}$ sorted in increasing order (again, when there are ties, items with lower index are ranked higher). To prove the lemma, it suffices to show that, when ϵ is sufficiently small,

$$\mathcal{R}(\mathbf{q}_\epsilon) \geq \mathcal{R}_\epsilon(\mathbf{p}_\epsilon) - O(n^3\epsilon).$$

To this end, we show that for any vector $\mathbf{u} \in V_\epsilon$ with $\mathbf{v} = \rho(\mathbf{u})$,

$$\mathcal{R}(\mathbf{v}, \mathbf{q}_\epsilon) \geq \mathcal{R}_\epsilon(\mathbf{u}, \mathbf{p}_\epsilon) - O(n^3\epsilon). \tag{11}$$

Finally we prove (11). First, we note that if $\mathcal{U}(\mathbf{v}, \widetilde{\mathbf{p}}) < 0$, then $\mathcal{R}(\mathbf{v}, \mathbf{q}_\epsilon) = \mathcal{R}_\epsilon(\mathbf{u}, \mathbf{p}_\epsilon) = 0$ when $\epsilon > 0$ is sufficiently small (as $\mathbf{u}$ approaches $\mathbf{v}$ and $\mathbf{p}_\epsilon$, $\mathbf{q}_\epsilon$ approach $\widetilde{\mathbf{p}}$). Otherwise, we have $\mathcal{U}(\mathbf{v}, \mathbf{q}_\epsilon) > \mathcal{U}(\mathbf{v}, \widetilde{\mathbf{p}}) \geq 0$ and we use k to denote the item selected in I on $(\mathbf{v}, \mathbf{q}_\epsilon)$. To violate (11), the item selected in I_ϵ on $(\mathbf{u}, \mathbf{p}_\epsilon)$ must be an item ℓ different from k satisfying $\widetilde{p}_\ell > \widetilde{p}_k$. Below we show that this cannot happen. Consider all the cases: 1) if $v_k - \widetilde{p}_k < v_\ell - \widetilde{p}_\ell$, we get a contradiction since item k is dominated by ℓ in I on $(\mathbf{v}, \mathbf{q}_\epsilon)$ when ϵ is sufficiently small; 2) if $v_k - \widetilde{p}_k > v_\ell - \widetilde{p}_\ell$, we get a contradiction with ℓ being selected in I_ϵ on $(\mathbf{u}, \mathbf{p}_\epsilon)$ when ϵ is sufficiently small; 3) if $v_k - \widetilde{p}_k = v_\ell - \widetilde{p}_\ell$ and $\widetilde{p}_\ell > \widetilde{p}_k$, we conclude that $v_k - q_{\epsilon,k} < v_\ell - q_{\epsilon,\ell}$, contradicting again with k being selected in I on $(\mathbf{v}, \mathbf{q}_\epsilon)$. (11) follows by combining all these cases. □

5. NP-hardness for support size 3

In this section we show:

Theorem 3. Item-Pricing-Decision *is NP-hard even when every distribution is supported on* $\{0, 1, 3\}$.

For this, we give a polynomial-time reduction from Partition to Item-Pricing-Decision for the case when distributions $\mathcal{F}_i$'s have support (at most) 3. Recall the Partition problem (Garey and Johnson, 1990):

Definition 5.1 *(*Partition*).*
Input: A set $C = \{c_1, \ldots, c_n\}$ of n positive integers (encoded in binary).
Problem: Does there exist a partition of C into two subsets with equal sum?

As Partition is NP-hard (Garey and Johnson, 1990), such a reduction implies that Item-Pricing-Decision is NP-hard.

Given an input instance of Partition, we construct an instance of Item-Pricing-Decision as follows. We have n items. Each item $i \in [n]$ can take 3 possible integer values $0, a, b$, where $b > a > 0$, i.e., $V_i = \{0, a, b\}$ for all i. Let $q_i = \Pr[v_i = b]$ and $r_i = \Pr[v_i = a]$. We set $q_i = c_i/M$ where $M = 2^n c_1^3$ and

$$r_i = \frac{b-a}{a(1-t_i)} \cdot q_i, \quad \text{where } t_i = \frac{b}{2a} \cdot \sum_{j \neq i, j \in [n]} q_j.$$

The two parameters a and b should be thought of as universal constants (independent of the given instance of Partition) throughout the proof. We will eventually set these constants to be $a = 1$, $b = 3$ (this choice is not necessary, there is flexibility in our proof and indeed any values with $b > 2a$ will work). However, for the sake of the presentation, we will keep a, b as generic parameters for most of the calculations till the end.

Note that the definition of r_i implies that

$$bq_i = a(q_i + r_i) - ar_it_i. \tag{12}$$

Let $N = 2^n c_1^2$. Then we have $q_i, r_i = O(1/N)$ and $t_i = O(n/N)$ for all i. *Thus, each distribution assigns most of its probability mass to the point* 0. This is a crucial property which allows us to get a handle on the optimal revenue. For an arbitrary general instance of the pricing problem, the expected revenue is a highly complex nonlinear function. The fact that most of the probability mass in our construction is concentrated at 0 implies that valuation vectors with many nonzero entries contribute very little to the expected revenue. As we will argue, the revenue is approximated well by its 1st and 2nd order terms with respect to poly$(n)/N$, which essentially corresponds to the contribution of all valuations in which at most two items have nonzero value. The probabilities q_i, r_i are chosen carefully so that the optimization of the expected revenue amounts to a quadratic optimization problem, which achieves its maximum possible value when the given set C of integers has a partition into two parts with equal sums.

Our main claim is that, for an appropriate value t^*, there exists a price vector with expected revenue at least t^* if and only if there exists a solution to the original instance of the Partition problem.

We start with some notation and simple observations. For $T_1, T_2, \epsilon \in \mathbb{R}$, we use $T_1 = T_2 \pm \epsilon$ to denote the inequality that $|T_1 - T_2| \leq \epsilon$. Note that, as both the q_i's and the t_i's are very small positive quantities, we have that $r_i \approx (b-a)q_i/a$. Formally, with the above notation we can write

$$r_i = \frac{b-a}{a(1-t_i)} \cdot q_i = \frac{b-a}{a} \cdot q_i \pm 2\frac{b-a}{a} \cdot q_it_i = \frac{b-a}{a} \cdot q_i \pm O(n/N^2). \tag{13}$$

Lemma 2.2 and Corollary 3.1 imply that a revenue maximizing price vector can be assumed to have non-negative integer coefficients of magnitude at most b. The following lemma establishes the stronger statement that, for our particular instance, an optimal price vector $\mathbf{p}$ can be assumed to have each p_i in the set $\{a, b\}$.

Lemma 5.1. *There is an optimal price vector* $\mathbf{p} \in \{a, b\}^n$.

Proof. By Lemma 2.2 and Corollary 3.1, there is an optimal price vector with integer coordinates in $[0 : b]$. Let $\mathbf{p}$ be any (integer) vector in $[0 : b]^n$ that has at least one coordinate $p_j \notin \{a, b\}$. We will show below that $\mathcal{R}(\mathbf{p}) < \mathcal{R}(\mathbf{b})$, where $\mathbf{b}$ denotes the all-b vector, and hence $\mathbf{p}$ is not optimal.

Consider an index $i \in [n]$ with $p_i > 0$. The probability the buyer selects item i is bounded from above by $\Pr[v_i \ge p_i]$, the probability that item i has value at least p_i, and is bounded from below by

$$\Pr\left[v_i \ge p_i\right] \cdot \prod_{j \ne i,\, j \in [n]} (1 - q_i - r_i) \ge \Pr\left[v_i \ge p_i\right] \cdot (1 - O(n/N)).$$

Note that the second term in the LHS above is the probability that all items other than i have value 0 and the inequality uses the fact that $q_i, r_i = O(1/N)$. Applying these two bounds on $\mathbf{p}$ and $\mathbf{b}$ we obtain

$$\mathcal{R}(\mathbf{b}) \ge \sum_{i \in [n]} q_i \,(1 - O(n/N)) \cdot b \quad \text{and} \quad \mathcal{R}(\mathbf{p}) \le \sum_{i : p_i > 0} \Pr\left[v_i \ge p_i\right] \cdot p_i.$$

So $\mathcal{R}(\mathbf{b}) \ge (\sum_{i \in [n]} q_i b) - O(n^2/N^2)$. Regarding $\mathcal{R}(\mathbf{p})$, we consider the following three cases. For $i \in [n]$ with $p_i = b$, the probability that $v_i \ge p_i$ is q_i and the contribution of such an item to the second sum is $q_i b$. Similarly, for $i \in [n]$ with $p_i = a$, the probability that $v_i \ge p_i$ is $q_i + r_i$ and the contribution to the sum is

$$(q_i + r_i)a \le q_i b + O(n/N^2),$$

where the inequality follows from (13). Finally, we consider an item $i \in [n]$ with $p_i \notin \{a, b\}$. If $a < p_i < b$ then the contribution is $q_i p_i$, which is at most $q_i(b-1) = q_i b - q_i$, since p_i is integer. If $p_i < a$, then the contribution is $(q_i + r_i)p_i$, which is at most $(q_i + r_i)(a - 1) = q_i b + a r_i t_i - q_i - r_i = q_i b - q_i - r_i(1 - a t_i)$. In both cases, the contribution to the sum is at most

$$q_i b - q_i \le q_i b - (1/M).$$

Note that the definition of M and N implies that $1/M \gg n^2/N^2$. Because there exists at least one j with $p_j \notin \{a, b\}$, it follows that $\mathcal{R}(\mathbf{p}) < \mathcal{R}(\mathbf{b})$ which completes the proof of the lemma. □

As a result, to maximize the expected revenue it suffices to consider price vectors in $\{a, b\}^n$. Given any price-vector $\mathbf{p} \in \{a, b\}^n$, we let $S = S(\mathbf{p}) = \{i \in [n] : p_i = a\}$ and $T = T(\mathbf{p}) = \{i \in [n] : p_i = b\}$. The main idea of the proof is to establish an appropriate *quadratic form approximation* to the expected revenue $\mathcal{R}(\mathbf{p})$ that is sufficiently accurate for the purposes of our reduction.

Approximating the revenue. We appropriately partition the valuation space V into three events that yield positive revenue. We then approximate the probability of each and its contribution to the expected revenue up to, and including, 2nd order terms, i.e., terms of order $O(\text{poly}(n)/N^2)$, and we ignore 3rd order terms, i.e., terms of order $O(\epsilon)$ where $\epsilon = n^3/N^3$.

In particular, we consider the following disjoint events:

- *First event*: $E_1 = \{\mathbf{v} \in V \mid \exists i \in S : v_i = b\}$.
  Note that for any $\mathbf{v} \in E_1$ we have $\mathcal{R}(\mathbf{v}, \mathbf{p}) = a$. The probability of this event is

$$\Pr[E_1] = 1 - \prod_{i \in S}(1 - q_i) = \sum_{i \in S} q_i - \sum_{i \ne j \in S} q_i q_j \pm O(\epsilon).$$

- *Second event*: $E_2 = \overline{E_1} \cap \{\mathbf{v} \in V \mid \exists i \in S : v_i = a \text{ and } \forall i \in T : v_i \in \{0, a\}\}$.
  Note that for any $\mathbf{v} \in E_2$ we have $\mathcal{R}(\mathbf{v}, \mathbf{p}) = a$. The probability of this event is

$$\Pr[E_2] = \prod_{j \in T}(1 - q_j) \left[\prod_{i \in S}(1 - q_i) - \prod_{i \in S}(1 - q_i - r_i) \right]$$

  Using the elementary identities

$$\prod_{j \in T}(1 - q_j) = 1 - \sum_{j \in T} q_j + \sum_{i \ne j \in T} q_i q_j \pm O(\epsilon)$$

$$\prod_{i \in S}(1 - q_i) = 1 - \sum_{i \in S} q_i + \sum_{i \ne j \in S} q_i q_j \pm O(\epsilon)$$

$$\prod_{i \in S}(1 - q_i - r_i) = 1 - \sum_{i \in S}(q_i + r_i) + \sum_{i \ne j \in S} (q_i + r_i)(q_j + r_j) \pm O(\epsilon),$$

we can write

$$\Pr[E_2] = \left[1 - \sum_{j\in T} q_j + \sum_{i\neq j\in T} q_i q_j \pm O(\epsilon)\right] \cdot \left[\sum_{i\in S} r_i + \sum_{i\neq j\in S} q_i q_j - \sum_{i\neq j\in S}(q_i + r_i)(q_j + r_j) \pm O(\epsilon)\right]$$
$$= \sum_{i\in S} r_i - \sum_{i\in S} r_i \sum_{j\in T} q_j + \sum_{i\neq j\in S} q_i q_j - \sum_{i\neq j\in S}(q_i + r_i)(q_j + r_j) \pm O(\epsilon).$$

- *Third event*: $E_3 = \overline{E_1} \cap \{\mathbf{v} \in V \mid \exists i \in T : v_i = b\}$.
  Note that for any $\mathbf{v} \in E_3$ we have $\mathcal{R}(\mathbf{v}, \mathbf{p}) = b$. The probability of this event is

$$\Pr[E_3] = \prod_{i\in S}(1 - q_i)\left[1 - \prod_{j\in T}(1 - q_j)\right]$$
$$= \left(1 - \sum_{i\in S} q_i + \sum_{i\neq j\in S} q_i q_j \pm O(\epsilon)\right)\left(\sum_{j\in T} q_j - \sum_{i\neq j\in T} q_i q_j \pm O(\epsilon)\right)$$
$$= \sum_{j\in T} q_j - \sum_{i\neq j\in T} q_i q_j - \sum_{i\in S} q_i \sum_{j\in T} q_j \pm O(\epsilon).$$

Therefore, for the expected revenue $\mathcal{R}(\mathbf{p})$ we have:

$$\mathcal{R}(\mathbf{p}) = \big(\Pr[E_1] + \Pr[E_2]\big) \cdot a + \Pr[E_3] \cdot b$$
$$= a \cdot \left(\sum_{i\in S}(q_i + r_i) - \sum_{i\neq j\in S}(q_i + r_i)(q_j + r_j) - \sum_{i\in S} r_i \sum_{j\in T} q_j\right)$$
$$+ b \cdot \left(\sum_{j\in T} q_j - \sum_{i\neq j\in T} q_i q_j - \sum_{i\in S} q_i \sum_{j\in T} q_j\right) \pm O(\epsilon).$$

Using (12) it follows that the first order term of the revenue is

$$b\sum_{j\in T} q_j + a\sum_{i\in S}(q_i + r_i) = b\sum_{j\in[n]} q_j + \sum_{i\in S}\big(a(q_i + r_i) - bq_i\big) = b\sum_{j\in[n]} q_j + \sum_{i\in S}(ar_i t_i).$$

Observe that the first term $b\sum_{j\in[n]} q_j$ in the above expression is a constant L_1, independent of the pricing (i.e., the partition of the items into S and T).

In the second order term, we can rewrite the expression $a\sum_{i\neq j\in S}(q_i + r_i)(q_j + r_j)$ as

$$\frac{1}{2} \cdot \sum_{i\in S}(q_i + r_i) \sum_{j\in S,\ j\neq i} a(q_j + r_j)$$
$$= \frac{1}{2} \cdot \sum_{i\in S}(q_i + r_i) \sum_{j\in S,\ j\neq i} (bq_j + ar_j t_j)$$
$$= \frac{b}{2} \cdot \sum_{i\in S} q_i \sum_{j\in S,\ j\neq i} q_j + \frac{b}{2} \cdot \sum_{i\in S} r_i \sum_{j\in S,\ j\neq i} q_j + \frac{1}{2} \cdot \sum_{i\in S}(q_i + r_i) \sum_{j\in S,\ j\neq i} ar_j t_j$$
$$= b\sum_{i\neq j\in S} q_i q_j + \frac{b}{2}\sum_{i\in S} r_i \sum_{j\in S,\ j\neq i} q_j \pm O(\epsilon)$$

where in the first expression above, the double summation is multiplied by $1/2$ because each unordered pair $i \neq j \in S$ is included twice. Thus, the second order term of the expected revenue $\mathcal{R}(\mathbf{p})$ is

$$-a\sum_{i\neq j\in S}(q_i + r_i)(q_j + r_j) - a\sum_{i\in S} r_i \sum_{j\in T} q_j - b\sum_{i\neq j\in T} q_i q_j - b\sum_{i\in S} q_i \sum_{j\in T} q_j$$
$$= -b\sum_{i\neq j\in S} q_i q_j - \frac{b}{2}\sum_{i\in S} r_i \sum_{j\in S,\ j\neq i} q_j - a\sum_{i\in S} r_i \sum_{j\in T} q_j - b\sum_{i\neq j\in T} q_i q_j - b\sum_{i\in S} q_i \sum_{j\in T} q_j \pm O(\epsilon)$$

$$= -b \sum_{i \neq j \in [n]} q_i q_j - \frac{b}{2} \sum_{i \in S} r_i \sum_{j \in S,\, \neq i} q_j - a \sum_{i \in S} r_i \sum_{j \in T} q_j \pm O(\epsilon).$$

The first term in the last expression is a constant L_2 independent of the pricing. As a result, we can rewrite the second order term as follows:

$$L_2 - \frac{b}{2} \sum_{i \in S} r_i \sum_{j \in S,\, j \neq i} q_j - a \sum_{i \in S} r_i \sum_{j \in T} q_j \pm O(\epsilon) = L_2 - \sum_{i \in S} r_i \left(\frac{b}{2} \sum_{j \in S,\, j \neq i} q_j + a \sum_{j \in T} q_j \right) \pm O(\epsilon).$$

Summing with the first order term and letting $L = L_1 + L_2$, we have:

$$\begin{aligned}
\mathcal{R}(\mathbf{p}) &= L + \sum_{i \in S} r_i \left(a t_i - \frac{b}{2} \sum_{j \in S,\, j \neq i} q_j - a \sum_{j \in T} q_j \right) \pm O(\epsilon) \\
&= L + \sum_{i \in S} r_i \left(\frac{b}{2} \sum_{j \neq i} q_j - \frac{b}{2} \sum_{j \in S,\, j \neq i} q_j - a \sum_{j \in T} q_j \right) \pm O(\epsilon) \\
&= L + \sum_{i \in S} r_i \cdot \left(\frac{b}{2} - a \right) \sum_{j \in T} q_j \pm O(\epsilon) \\
&= L + \frac{b-a}{a} \cdot \left(\frac{b}{2} - a \right) \cdot \frac{1}{M^2} \cdot \sum_{i \in S} c_i \cdot \sum_{j \in T} c_j \pm O(\epsilon).
\end{aligned}$$

Now setting $a = 1, b = 3$ in the previous expression, we have that for any $\mathbf{p} \in \{a, b\}^n$,

$$\mathcal{R}(\mathbf{p}) = L + \frac{1}{M^2} \left(\sum_{i \in S} c_i \right) \cdot \left(\sum_{j \in T} c_j \right) \pm O(\epsilon). \tag{14}$$

At this point, we observe that the sum of the two factors $\sum_{i \in S} c_i, \sum_{j \in T} c_j$ in (14) is a constant (independent of the partition). Thus, their product is maximized when they are equal. Because $\epsilon = o(1/M^2)$, it follows that the revenue is maximized when the product of the two factors is maximized. In particular, if there exists a partition of the set $C = \{c_1, \ldots, c_n\}$ into two sets with equal sums $H = (\sum_{i \in [n]} c_i)/2$, then the corresponding partition of the indices into the sets S and T yields revenue

$$L + \frac{1}{M^2} \cdot H^2 \pm O(\epsilon).$$

On the other hand, if there is no such equipartition of C, then for any partition, the revenue will be at most

$$L + \frac{1}{M^2} (H+1)(H-1) \pm O(\epsilon) = L + \frac{1}{M^2} (H^2 - 1) \pm O(\epsilon).$$

Since $\epsilon = o(1/M^2)$ it follows that there exists a partition of the set $C = \{c_1, \ldots, c_n\}$ into two sets with equal sums if and only if there exists a price vector $\mathbf{p} \in \{a, b\}^n$ with

$$\mathcal{R}(\mathbf{p}) \geq t^* = L + \frac{1}{M^2} \left(H^2 - \frac{1}{2} \right).$$

This completes the proof. $\square$

Remark. In the above construction, the support $\{0, a, b\}$ of the distributions includes the value 0 (which in fact has most of the probability mass). It is easy to modify the construction, if desired, so that the support contains only positive values: shift all the values of the distributions up by 1 (thus, the supports now become $V_i = \{1, 2, 4\}$) and add an additional $(n+1)$-th item which has value 1 with probability 1. This transformation increases the expected revenue by 1. It is easy to see that an optimal price vector $\mathbf{p}'$ for the new instance will give price $p'_{n+1} = 1$ to the $(n+1)$-th item and price $p'_i = p_i + 1$ to each other item $i \in [n]$, where $\mathbf{p}$ is an optimal vector for the original instance.

6. NP-hardness for identical distributions

In this section we show:

Theorem 4. Item-Pricing-Decision *is NP-hard even when the distributions are identical.*

For this purpose we reduce from the following (still NP-complete) version of Integer Knapsack.

Definition 6.1 *(*Integer Knapsack with Repetitions*).*
Input: $n+1$ positive integers $a_1 < \cdots < a_n$ and L.
Problem: Do there exist nonnegative integers $x_1, \ldots, x_n$ such that $\sum_{i\in[n]} x_i = n$ and $\sum_{i\in[n]} x_i a_i = L$?

The NP-hardness of this version of Integer Knapsack is likely known in the literature. For completeness we include a quick proof via a reduction from Subset-Sum, a classical NP-complete problem (Garey and Johnson, 1990):

Definition 6.2 *(*Subset-Sum*).*
Input: $n+1$ positive integers $b_1 < \cdots < b_n$ and T.
Problem: Does there exist a subset S of $\{b_1, \ldots, b_n\}$ such that $\sum_{b_i\in S} b_i = T$?

Lemma 6.1. Integer Knapsack with Repetitions *is NP-hard.*

Proof. Let $b_1 < \cdots < b_n$ and T denote an instance of Subset-Sum, where b_i and T are all positive integers. Without loss of generality, we assume that $T > b_n$. Let $K = n^2 T$. For each $i \in [n]$, set $a_i = K^i + b_i$ and $c_i = K^i$. Then one can see that $\{K^{n+1}, a_i, c_i : i \in [n]\}$, a set of $2n+1$ positive integers, together with

$$L = T + K + K^2 + \cdots + K^n + (n+1)K^{n+1}$$

form a yes-instance of the special Integer Knapsack problem iff a subset of $\{b_1, \ldots, b_n\}$ sums to T. $\square$

Using Lemma 6.1, a polynomial-time reduction from the Integer Knapsack with Repetitions problem to our Item-Pricing-Decision would imply that the latter is NP-hard as well.

6.1. Intuition

The reason why we choose to reduce from Integer Knapsack with Repetitions instead of Partition or Subset-Sum is as follows. While a solution to $\{a_1, \ldots, a_n\}$ and L is formulated as n integers $x_1, \ldots, x_n$ that sum to n and satisfy $\sum_i x_i a_i = L$, one can equivalently view it as picking n integers from $\{a_1, \ldots, a_n\}$ with repetitions round by round, for n rounds. Notably the set of possible actions one can choose in each round is exactly the same, i.e., picking an integer from $\{a_1, \ldots, a_n\}$. This helps build a connection to the i.i.d. item pricing problem since one can imagine to use each item, with the same probability distribution, to mimic the action of choosing an integer from $\{a_1, \ldots, a_n\}$. This feature of Integer Knapsack with Repetitions is not shared by Subset-Sum (or Partition), where one can similarly view a solution as a sequence of n actions but in the ith round one decides whether the ith integer is included in the subset or not.

Let $a_1 < \cdots < a_n$ and L denote an instance of Integer Knapsack with Repetitions. Without loss of generality, we can assume that $L \le na_n$; otherwise the problem is trivial. Our goal is to construct a distribution Q over nonnegative integers, and reduce the problem Integer Knapsack with Repetitions to Item-Pricing-Decision with n items, each of which has its value drawn from Q independently. The key idea is similar to the reduction for support size 3 in Section 5, which we sketch below, but its implementation is more challenging.

At a high level, Q is supported on $\{0, v_1, \ldots, v_n\}$, where $0 < v_1 < \cdots < v_n$ are integers and grow exponentially. Q assigns most of its probability mass to the point 0, so that valuations with at least three positive entries contribute very little to the expected revenue. The hope is that we can set probabilities of $v_1, \ldots, v_n$ so that

1. There is an optimal price vector $\mathbf{p} \in \{v_1, \ldots, v_n\}^n$;
2. Let x_i denote the number of items priced at v_i in a price vector $\mathbf{p} \in \{v_1, \ldots, v_n\}^n$. Then the expected revenue obtained by $\mathbf{p}$ can be well approximated by the following quadratic form:

$$C_1 - C_2 \cdot \left(\sum_{i\in[n]} x_i a_i - L \right)^2, \tag{15}$$

for some positive constants C_1 and C_2 that do not depend on $\mathbf{p}$. It would then follow from similar arguments used in Section 5 that, given the optimal expected revenue of the i.i.d. instance, one can easily tell whether $\{a_1, \ldots, a_n\}$ and L is a yes-instance of Integer Knapsack with Repetitions or not.

Note that the plan is essentially the same as that of Section 5, except that (1) each item has n candidate prices instead of 2; (2) the distributions are i.i.d. and each mimics the action of picking one integer from $\{a_1, \ldots, a_n\}$, while each item i has a separate distribution in Section 5 and mimics the allocation of the ith integer in the partition.

The most challenging part of the construction is to make sure that the expected revenue indeed takes a form as in (15). For this purpose Q will actually have a significantly larger support than $\{0, v_1, \ldots, v_n\}$; it will be supported on

$$\Big\{0, v_i, v_i + j : i \in [n] \text{ and } j \in [2n^3]\Big\}. \tag{16}$$

The gaps between v_i's are significantly larger than n^3 so the j-part should be viewed as a small perturbation on v_i.

The proof is organized as follows. In Section 6.2 we formally construct the distribution Q from $\{a_1, \ldots, a_n\}$ and L. Scales of probabilities in Q will be fixed but their exact values will be specified later in the proof. We then show in Lemma 6.2 that there exists an optimal price vector $\mathbf{p} \in \{v_1, \ldots, v_n\}^n$. Next in Section 6.3 we analyze and derive an approximation of the expected revenue obtained by a price vector $\mathbf{p} \in \{v_1, \ldots, v_n\}^n$. The rest of the section is devoted to a long process of reverse engineering on exact values of probabilities in Q so that the expected revenue obtained by $\mathbf{p}$ can be written as (15) for appropriate constants C_1 and C_2.

6.2. Reduction

We start the construction of Q with some parameters. Let $m = \max(n^5, a_n)$, and let $N = m^{n^2}$ denote a large integer. For each $i \in [n]$, let $v_i = m^{n+i}$. For each $i \in [n-1]$, let

$$\gamma_i = \frac{1}{N}\left(\frac{1}{m^{n+i}} - \frac{1}{m^{n+i+1}}\right) = \frac{m-1}{Nm^{n+i+1}}.$$

Let $\gamma_n = 1/(Nm^{2n})$. For convenience, we also let $\Gamma_i = \sum_{j=i}^{n} \gamma_j = 1/(Nm^{n+i})$ for each $i \in [n]$.

We record a property that follows directly from our choices of v_i and γ_i.

Property 6.1. *For each $i \in [n]$, we have $v_i \Gamma_i = 1/N$.*

Let $\mathbf{q}_1, \ldots, \mathbf{q}_n$ denote n probability distributions. They are closely related to the instance of the Integer Knapsack problem and will be specified later in this section. The support of each $\mathbf{q}_i$ is a subset of $[2n^3]$ and for each $j \in [2n^3]$, we use $q_i(j)$ to denote the probability of j in $\mathbf{q}_i$. Finally, let $t_1, \ldots, t_n$ denote a sequence of (not necessarily positive) numbers, also to be specified later, with $|t_i| = O(1/N^2)$ for all $i \in [n]$.

We are ready to define Q using v_i, γ_i, t_i and $\mathbf{q}_i$. First the support of Q is given in (16). Note that all values in the support are bounded by $O(m^{2n})$, and the size of the support is $O(n^4)$.

Next Q has probability $(\gamma_i/m) + t_i$ at v_i for each $i \in [n]$; probability $q_i(j) \cdot \gamma_i(m-1)/m$ at $v_i + j$ for each $i \in [n]$ and $j \in [2n^3]$; and probability $1 - (\sum_{i=1}^{n} \gamma_i + t_i)$ at 0. It is easy to verify that Q is a probability distribution since the probabilities sum to 1. For convenience, we also let

$$T_i = \sum_{j=i}^{n} t_j \quad \text{and} \quad r_i = \sum_{j=i}^{n} (\gamma_j + t_j) = \Gamma_i + T_i,$$

for each $i \in [n]$. The latter quantity, r_i, is the probability that the value is at least v_i.

Even though t_i and $\mathbf{q}_i$ have not been specified yet, we still can prove the following useful lemma about optimal price vectors, as long as $|t_i| = O(1/N^2)$ for each $i \in [n]$:

Lemma 6.2. *There is an optimal price vector $\mathbf{p} \in \{v_1, \ldots, v_n\}^n$.*

Proof. By Lemma 2.2 and Corollary 3.1 there must be an (integral) optimal price vector in $[0 : v_n + 2n^3]^n$.

Let $\mathbf{p} = (p_1, \ldots, p_n) \in [0 : v_n + 2n^3]^n$ be a price vector with $\mathbf{p} \notin \{v_1, \ldots, v_n\}^n$. We will prove below that $\mathcal{R}(\mathbf{p}) < \mathcal{R}(\mathbf{b})$, where $\mathbf{b}$ is the vector in which all entries are v_n. The lemma then follows.

For convenience, we use $F(s)$ to denote the probability of a random variable drawn from the distribution Q being at least s. For each index $i \in [n]$ such that $p_i > 0$, the probability that the buyer picks item i can be bounded from above by $F(p_i)$, and can be bounded from below by

$$F(p_i) \cdot (1 - r_1)^{n-1} \ge F(p_i) \cdot \left(1 - \frac{1}{m^{n+1}N} - O\left(\frac{n}{N^2}\right)\right)^{n-1} \ge F(p_i) - O\left(\frac{n}{m^{2n+2}N^2}\right),$$

where we used $r_1 = \Gamma_1 + T_1$, $\Gamma_1 = 1/(m^{n+1}N)$, $T_1 = O(n/N^2)$ and $F(p_i) \le r_1 = O(1/(m^{n+1}N))$ if $p_i > 0$. Applying the upper bound on $\mathcal{R}(\mathbf{p})$ and the lower bound on $\mathcal{R}(\mathbf{b})$, we have

$$\mathcal{R}(\mathbf{p}) \le \sum_{i:p_i>0} F(p_i)\cdot p_i \quad \text{and} \quad \mathcal{R}(\mathbf{b}) \ge n v_n \left(F(v_n) - O\left(\frac{n}{m^{2n+2}N^2}\right)\right) \ge n v_n F(v_n) - O\left(\frac{n^2}{m^2N^2}\right).$$

We now examine $p_iF(p_i)$ and $v_nF(v_n)$. We have three cases on $sF(s)$:

Case 1: $s = v_i$ for some $i \in [n]$. Then we have

$$sF(s) = v_i(\Gamma_i + T_i) = \frac{1}{N} \pm O\left(\frac{nm^{2n}}{N^2}\right).$$

Case 2: $s = v_i + j$ for some $i \in [n]$ and $j \in [2n^3]$. We then have $F(s) \le r_i - (\gamma_i/m) - t_i$ and

$$sF(s) \le (v_i + 2n^3)\left(r_i - \frac{\gamma_i}{m} - t_i\right) = \frac{1}{N}\cdot\frac{m^2-m+1}{m^2} + O\left(\frac{n^3}{m^{n+1}N}\right) = \frac{1}{N} - \Omega\left(\frac{1}{mN}\right)$$

when $i < n$, and similarly when $i = n$,

$$sF(s) \le (v_n + 2n^3)\cdot\frac{\gamma_n(m-1)}{m} = \frac{m-1}{m}\cdot\frac{1}{N} + O\left(\frac{n^3}{m^{2n}N}\right) = \frac{1}{N} - \Omega\left(\frac{1}{mN}\right).$$

Case 3: Otherwise, let $i \in [n]$ denote the smallest index such that $s < v_i$. Then we have

$$sF(s) \le (v_i - 1)r_i = v_i(\Gamma_i + T_i) - r_i = \frac{1}{N} - \Omega\left(\frac{1}{m^{2n}N}\right).$$

From Case 1, we have

$$\mathcal{R}(\mathbf{b}) \ge \frac{n}{N} - O\left(\frac{n^2m^{2n}}{N^2}\right).$$

Regarding $\mathcal{R}(\mathbf{p})$, combining all three cases, we have that

$$\mathcal{R}(\mathbf{p}) \le \frac{n}{N} - \Omega\left(\frac{1}{m^{2n}N}\right)$$

because there is at least one index $i \in [n]$ such that $p_i \notin \{v_1, \dots, v_n\}$ by the assumption. As $N \gg n^2m^{4n}$, we conclude that $\mathcal{R}(\mathbf{p}) < \mathcal{R}(\mathbf{b})$. The lemma then follows. □

6.3. Analysis of the expected revenue

Given a price vector $\mathbf{p} \in \{v_1, \dots, v_n\}^n$, we let x_i denote the number of items priced at v_i. Then $\sum_i x_i = n$. We will only consider the contribution of two types of valuation vectors to the expected revenue $\mathcal{R}(\mathbf{p})$: those with exactly one positive entry and those with exactly two positive entries. The following lemma shows that the total contribution from all other valuation vectors is of third order with respect to (roughly) $1/N$.

Lemma 6.3. *The revenue from valuation vectors with at least three positive entries is $O(n^3/(m^{n+3}N^3))$.*

Proof. The probability that a valuation vector has at least three positive entries can be bounded by

$$O(n^3r_1^3) = O\left(\frac{n^3}{m^{3n+3}N^3}\right).$$

Thus, the total contribution is at most $O(m^{2n})\cdot O(n^3r_1^3)$, and the lemma follows. □

Let $\epsilon = n^3/(m^{n+3}N^3)$ and $\epsilon' = n^3m^{n-1}/N^3$. In the rest of Section 6.3, we show that the following explicit expression is an approximation of the expected revenue $\mathcal{R}(\mathbf{p})$ with error $O(\epsilon')$:

$$\frac{n}{N} + \sum_{i\in[n]} x_iv_iT_i - \frac{n-1}{2N}\sum_{i\in[n]} x_ir_i + \sum_{i<j\in[n]}\frac{x_ix_j}{N}\cdot\big((1/2) - p(i,j)\big)(\Gamma_i - \Gamma_j), \tag{17}$$

where, for each pair $i < j \in [n]$, we use $p(i,j) \in [0,1]$ to denote the probability of $\alpha - v_i > \beta - v_j$, where α and β are drawn independently from Q conditioning on $\alpha \ge v_i$ and $\beta \ge v_j$. This is done by examining closely valuations vectors with either one or two positive entries and performing some detailed calculations to simplify their contribution to the expected revenue. Once the approximation (17) is obtained, we show in the rest of the section that the t_i's and $\mathbf{q}_i$'s can be chosen carefully so that (17) can be further simplified to an expression of the form $c_1 - c_2 \sum_{i\in[n]}(x_i a_i - L)^2$, where

$$c_1 = \frac{n}{N} + \frac{L^2}{N^2 m^{3n}} \quad \text{and} \quad c_2 = \frac{1}{N^2 m^{3n}}$$

are positive constants that depend only on n, m, N and L. Note that this expression attains its maximum value c_1 when $\sum_{i\in[n]} x_i a_i = L$, i.e. when the given input instance of Integer Knapsack with Repetitions has a solution, and this fact is used to finish the reduction.

We start with valuation vectors with exactly one positive entry. Their total contribution is

$$\sum_{i\in[n]} x_i v_i r_i (1-r_1)^{n-1}.$$

Since $r_1 = O(1/(m^{n+1}N))$ is of first order, approximating the sum up to second order yields

$$\begin{aligned}\sum_{i\in[n]} x_i v_i r_i (1-r_1)^{n-1} &= \sum_{i\in[n]} x_i v_i r_i \big(1-(n-1)r_1 \pm O(n^2 r_1^2)\big)\\ &= \sum_{i\in[n]} x_i v_i r_i - (n-1)\sum_{i\in[n]} x_i v_i r_i r_1 \pm O(\epsilon). \end{aligned} \tag{18}$$

The contribution of valuation vectors with two positive entries is more involved. First from those whose two positive entries are over items of the same price, the total contribution to $\mathcal{R}(\mathbf{p})$ is

$$\sum_{i\in[n]} \frac{x_i(x_i-1)}{2} \cdot v_i\big(r_1^2 - (r_1-r_i)^2\big)(1-r_1)^{n-2}. \tag{19}$$

Using $p(i,j)$'s, the contribution from vectors whose two positive entries are on items of different prices is

$$\sum_{i<j\in[n]} x_i x_j \Big(v_i r_i (r_1 - r_j) + v_j r_j (r_1 - r_i) + r_i r_j \Big(v_i p(i,j) + v_j\big(1-p(i,j)\big)\Big)\Big)(1-r_1)^{n-2}. \tag{20}$$

Approximating to the second order, (19) can be simplified to

$$\sum_{i\in[n]} \frac{x_i(x_i-1)}{2} \cdot v_i(2r_i r_1 - r_i^2)(1 \pm O(nr_1)) = \sum_{i\in[n]} x_i(x_i-1) v_i r_i r_1 - \sum_{i\in[n]} \frac{x_i(x_i-1)}{2} \cdot v_i r_i^2 \pm O(\epsilon) \tag{21}$$

and (20) can be simplified similarly to

$$\sum_{i<j\in[n]} x_i x_j \Big(v_i r_i (r_1 - r_j) + v_j r_j (r_1 - r_i) + r_i r_j \Big(v_i p(i,j) + v_j\big(1-p(i,j)\big)\Big)\Big) \pm O(\epsilon). \tag{22}$$

Next we show that, for each $i \in [n]$, all terms of $v_i r_i r_1$ in (18), (21) and (22) cancel each other. This is because the overall coefficient of $v_i r_i r_1$ is

$$-(n-1)x_i + x_i(x_i-1) + \sum_{j:j\neq i} x_i x_j = -(n-1)x_i + nx_i - x_i = 0,$$

where the first equality uses the fact that $\sum_{j\in[n]} x_j = n$. This allows us to further simplify the sum of (18), (21) and (22), with an error of $O(\epsilon)$, to

$$\sum_{i\in[n]} x_i v_i r_i - \sum_{i\in[n]} \frac{x_i(x_i-1)}{2} \cdot v_i r_i^2 - \sum_{i<j\in[n]} x_i x_j r_i r_j (v_i + v_j) + \sum_{i<j\in[n]} x_i x_j r_i r_j \big(v_i p(i,j) + v_j(1-p(i,j))\big). \tag{23}$$

Note that, by Lemma 6.3, this is also an approximation of $\mathcal{R}(\mathbf{p})$, with an error of $O(\epsilon)$.

Recall that $\epsilon' = n^3 m^{n-1}/N^3$. Plugging in $v_i r_i = v_i(\Gamma_i + T_i) = (1/N) + v_i T_i$ (note that $T_i = O(n/N^2)$ is of second order), (23) can be further simplified to the following:

$$\frac{n}{N} + \sum_{i\in[n]} x_i v_i T_i - \sum_{i\in[n]} \frac{x_i r_i (x_i - 1)}{2N} - \sum_{i<j\in[n]} \frac{x_i x_j (r_i + r_j)}{N} + \sum_{i<j\in[n]} \frac{x_i x_j}{N} \cdot \big(r_j p(i,j) + r_i(1-p(i,j))\big) \pm O(\epsilon').$$

Extracting $x_i x_j (r_i + r_j)/(2N)$ from the last sum above, we get

$$\frac{n}{N} + \sum_{i\in[n]} x_i v_i T_i - \sum_{i\in[n]} \frac{x_i r_i (x_i - 1)}{2N} - \sum_{i<j\in[n]} \frac{x_i x_j (r_i + r_j)}{2N} + \sum_{i<j\in[n]} \frac{x_i x_j}{N} \cdot \big((1/2) - p(i,j)\big)(r_i - r_j) \pm O(\epsilon').$$

Also note that the second and third sums above can be combined into a linear form of the x_i's:

$$\sum_{i\in[n]} x_i r_i (x_i - 1) + \sum_{i<j\in[n]} x_i x_j (r_i + r_j) = -\sum_{i\in[n]} x_i r_i + \left(\sum_{i\in[n]} x_i\right)\left(\sum_{i\in[n]} x_i r_i\right) = (n-1)\sum_{i\in[n]} x_i r_i.$$

As a result, we finally get (17) as an approximation of the expected revenue $\mathcal{R}(\mathbf{p})$, with an error of $O(\epsilon')$. (Note that in (17) we also replaced $r_i - r_j$ at the end with $\Gamma_i - \Gamma_j$ since the error introduced is $O(n^3/N^3)$.)

6.4. Reverse engineering of t_i and $p(i, j)$

Our ultimate goal is to set t_i's and $\mathbf{q}_i$'s carefully so that (17) by the end has the following form:

$$\frac{n}{N} + \frac{L^2}{N^2 m^{3n}} - \frac{1}{N^2 m^{3n}} \cdot \left(\sum_{i\in[n]} x_i a_i - L\right)^2. \tag{24}$$

Recall that L is the target in the Integer Knapsack instance. If this is the case, we obtain a polynomial-time reduction from the Integer Knapsack with Repetitions to Item-Pricing-Decision, because the difference between (24) and $\mathcal{R}(\mathbf{p})$ is at most $O(\epsilon')$ and thus (24) is at least

$$\frac{n}{N} + \frac{L^2}{N^2 m^{3n}} - \frac{1}{2N^2 m^{3n}}$$

if and only if $a_1, \ldots, a_n$ and L is a yes-instance of Integer Knapsack with Repetitions.

To compare (24) and (17), we use $\sum_{i\in[n]} x_i = n$ in (24) and it becomes

$$\begin{aligned}
&\frac{n}{N} - \frac{1}{N^2 m^{3n}} \cdot \left(\sum_{i\in[n]} x_i^2 a_i^2 + 2\sum_{i<j\in[n]} x_i x_j a_i a_j - 2\sum_{i\in[n]} a_i L x_i\right)\\
&= \frac{n}{N} - \frac{1}{N^2 m^{3n}} \cdot \left(\sum_{i\in[n]} a_i^2 x_i \left(n - \sum_{j:j\neq i} x_j\right) + 2\sum_{i<j\in[n]} x_i x_j a_i a_j - 2\sum_{i\in[n]} a_i L x_i\right)\\
&= \frac{n}{N} - \frac{1}{N^2 m^{3n}} \cdot \left(\sum_{i\in[n]} (n a_i^2 - 2a_i L) x_i - \sum_{i<j\in[n]} x_i x_j (a_i - a_j)^2\right).
\end{aligned} \tag{25}$$

By comparing (25) with (17), our goal is achieved if the following two conditions hold: First,

$$T_i = \frac{1}{v_i} \cdot \left(\frac{(n-1) r_i}{2N} - \frac{1}{N^2 m^{3n}} \cdot \left(n a_i^2 - 2a_i L\right)\right), \tag{26}$$

for all $i \in [n]$ (note that the absolute value of the right side of (26) is $O(n/(m^{2n+2}N^2))$; Second,

$$\frac{((1/2) - p(i,j))(\Gamma_i - \Gamma_j)}{N} = \frac{(a_i - a_j)^2}{N^2 m^{3n}}, \quad \text{for all pairs } i < j \in [n]. \tag{27}$$

For the first condition, we note that the equations (26) for all $i \in [n]$ actually form a triangular system of n equations in the n variables $t_1, \ldots, t_n$, and thus there exists a unique sequence $t_1, \ldots, t_n$ such that (26) holds for all $i \in [n]$. Moreover, as the absolute value of the right side of (26) is $O(n/(m^{2n+2}N^2))$, the t_i's are $O(1/N^2)$ as we promised earlier. To see this, we let s denote the maximum of the absolute value of the right side of (26), over all $i \in [n]$. Then one can show by induction on i that $|t_i| \leq 2^{n-i}s$ for all i from n to 1. The claim now follows using $2^n \ll m^n$.

The second condition is more difficult to satisfy. From (27), we know that the condition is met if

$$\frac{1}{2} - p(i,j) = \frac{(a_i - a_j)^2}{N m^{3n} (\Gamma_i - \Gamma_j)}, \quad \text{for all } i < j \in [n]. \tag{28}$$

We will define below the n distributions $\mathbf{q}_i$, $i \in [n]$, so that their induced values for the probabilities $p(i, j)$ satisfy (28). An important property that we will need for the construction of the $\mathbf{q}_i$'s is that all the desired probabilities $p(i, j)$ are very close to 1/2. Specifically, using $\Gamma_i - \Gamma_j \geq \gamma_i \geq \gamma_n = 1/(m^{2n} N)$, we have

$$0 < \frac{1}{2} - p(i,j) \le \frac{(a_i - a_j)^2 \cdot Nm^{2n}}{Nm^{3n}} = o\left(\frac{1}{m}\right), \tag{29}$$

since $m = \max(n^5, a_n)$ and $a_n = \max_{i\in[n]} a_i$.

6.5. Connecting $p(i,j)$ with $\mathbf{q}_i$ and $\mathbf{q}_j$

Fixing a pair $i < j \in [n]$, we examine $p(i,j)$ closer. Recall that $p(i,j)$ is the probability of $\alpha - v_i > \beta - v_j$ when α and β are drawn independently from Q, conditioning on $\alpha \ge v_i$ and $\beta \ge v_j$.

For convenience, we use *block* k to denote the subset $\{v_k, v_k+1, \dots, v_k+2n^3\}$ of the support of Q. Note that due to the exponential structure of the support of Q (and the assumption of $i<j$), if α is in block $k \ge i$ and β is in block $\ell > j$ with $\ell > k$ then $\beta - v_j > \alpha - v_i$. Therefore, for $\alpha - v_i > \beta - v_j$ to happen, we only need to consider the following three cases:

Case 1: α is from block k and β is from block ℓ, where $k,\ell \in [n]$ satisfy $k \ge \ell > j$. Then the total contribution of this case to probability $p(i,j)$ is:

$$\frac{1}{r_i r_j} \cdot \sum_{k \ge \ell > j} (\gamma_k + t_k)(\gamma_\ell + t_\ell).$$

Case 2: α is from block k and β is from block j, where $k > i$. Then the total contribution is

$$\frac{1}{r_i r_j} \cdot \sum_{k>i} (\gamma_k + t_k)(\gamma_j + t_j).$$

Case 3: Finally, α is from block i and β is from block j, with $\alpha - v_i > \beta - v_j$. Let $q(i,j)$ denote the probability of $\alpha > \beta$, when α is drawn from $\mathbf{q}_i$ and β is drawn from $\mathbf{q}_j$ independently. Using $q(i,j)$, the total contribution of this case to $p(i,j)$ is

$$\frac{1}{r_i r_j} \cdot \left(\left(\frac{\gamma_j}{m} + t_j\right) \cdot \frac{(m-1)\gamma_i}{m} + q(i,j) \cdot \frac{(m-1)\gamma_i}{m} \cdot \frac{(m-1)\gamma_j}{m} \right).$$

The probability $p(i,j)$ is equal to the sum of the above three quantities for the three cases. Hence, $q(i,j)$ is uniquely determined by the $p(i,j)$ we aim for, i.e., the unique $p(i,j)$ that satisfies (28), because all other parameters have been well defined by now, including $t_1, \dots, t_n$.

We show below that, if $|p(i,j) - 1/2| = o(1/m)$, then $q(i,j)$ must satisfy $|q(i,j) - 1/2| = O(1/m)$.

To see this, note first that since $i < j \le n$ and $r_i = \Gamma_i + T_i = 1/(Nm^{n+i}) \pm O(n/N^2)$, we have that

$$\gamma_i = \frac{m-1}{m^{n+i+1}N} = \frac{m-1}{m} \cdot r_i \pm O\left(\frac{n}{N^2}\right).$$

Thus, $\sum_{k>i} (\gamma_k + t_k) = r_i - \gamma_i - t_i = r_i/m \pm O(n/N^2)$.

Using this fact in the above expressions for the three cases, it is easy to show that, other than

$$\frac{1}{r_i r_j} \cdot q(i,j) \cdot \frac{(m-1)\gamma_i}{m} \cdot \frac{(m-1)\gamma_j}{m}, \tag{30}$$

the contribution of other terms is bounded from above by $O(1/m)$ (note that $k \ge \ell > j$ implies $k > i$). Since $|p(i,j) - 1/2| = o(1/m)$, it follows that the term in (30) is between $1/2 - O(1/m)$ and $1/2 + O(1/m)$. Note that $\gamma_i = (m-1)r_i/m \pm O(n/N^2)$ (since $i<n$), and γ_j is either $(m-1)r_j/m \pm O(n/N^2)$ if $j<n$ or $r_j \pm O(n/N^2)$ if $j=n$. Therefore, the coefficient of $q(i,j)$ in (30) is $1 - O(1/m)$. Since the expression in (30) is $1/2 \pm O(1/m)$, it follows that $|q(i,j) - 1/2| = O(1/m)$.

6.6. Reverse engineering of $\mathbf{q}_i$

Given $q(i,j)$ for each pair $i<j \in [n]$, our final technical step of the reduction is to construct a sequence of probability distributions $\mathbf{q}_1, \dots, \mathbf{q}_n$ over $[2n^3]$ such that, for each pair $i<j\in[n]$, the probability of $\alpha > \beta$, where α is drawn from $\mathbf{q}_i$ and β is drawn from $\mathbf{q}_j$ independently, is exactly $q(i,j)$.

In general, such a sequence of distributions may not exist, e.g., consider $n=3$, $q(1,2)=1$, $q(2,3)=1$ and $q(1,3)=0$. But here we are guaranteed that the $q(i,j)$'s are close to 1/2: $|q(i,j) - 1/2| = O(1/m)$. We shall show that in this case the desired distributions exist, and we can construct them.

To construct $\mathbf{q}_1, \dots, \mathbf{q}_n$, we define $\binom{n}{2}$ subsets of $[2n^3]$, called *sections*. Each section consists of $2n+3$ consecutive integers. The first section is $\{1, \dots, 2n+3\}$, the second section is $\{2n+4, \dots, 4n+6\}$, and so on and so forth. (Note that $2n^3$ is clearly large enough for $\binom{n}{2}$ sections.) Each section is labeled, arbitrarily, by a distinct pair (i,j) with $i<j\in[n]$. We let $t_{i,j,k}$ denote the kth smallest integer in section (labeled) (i,j), where $k \in [2n+3]$. Now we define $\mathbf{q}_\ell$, $\ell \in [n]$. For each section (i,j), $i<j\in[n]$, we have:

Case 1: If $\ell \neq i$ and $\ell \neq j$, then we set

$$q_\ell\big(t_{i,j,\ell}\big) = q_\ell\big(t_{i,j,2n+4-\ell}\big) = \frac{1}{2\binom{n}{2}}$$

and $q_\ell\big(t_{i,j,k}\big) = 0$ for all other $k \in [2n+3]$.

Case 2: If $\ell = j$, then we set $q_\ell\big(t_{i,j,n+2}\big) = 1/\binom{n}{2}$ and $q_\ell\big(t_{i,j,k}\big) = 0$ for all other $k \in [2n+3]$.

Case 3: If $\ell = i$, then we set

$$q_\ell\big(t_{i,j,n+1}\big) = \frac{1}{2\binom{n}{2}} - \binom{n}{2}\big(q(i,j) - 1/2\big) \quad \text{and} \quad q_\ell\big(t_{i,j,n+3}\big) = \frac{1}{2\binom{n}{2}} + \binom{n}{2}\big(q(i,j) - 1/2\big),$$

and $q_\ell\big(t_{i,j,k}\big) = 0$ for all other $k \in [2n+3]$.

This finishes the construction of $\mathbf{q}_1, \dots, \mathbf{q}_n$. Using $|q(i,j) - 1/2| = O(1/m)$ and $m \geq n^5$, we know that $\mathbf{q}_1, \dots, \mathbf{q}_n$ are probability distributions: all entries are nonnegative and sum to 1.

It is also not hard to verify that the distributions satisfy the desired property, i.e., for each pair $i < j \in [n]$ the probability of $\alpha > \beta$, where α is drawn from $\mathbf{q}_i$ and β is drawn from $\mathbf{q}_j$ independently, is exactly $q(i,j)$. First observe that every section of each distribution $\mathbf{q}_i$ has the same probability $1/\binom{n}{2}$. If α and β belong to different sections then the order between α and β is determined by the order of the sections, and both orders have obviously the same probability.

So suppose that α, β belong to the same section labeled (g,h), where $g, h \in [2n+3]$. If $g \neq i$ or $h \neq j$, then it is easy to check that both orders between α and β have the same probability. To see this, suppose first that $i \notin \{g,h\}$. Then $\alpha = t_{g,h,i}$ or $t_{g,h,2n+4-i}$ with equal probability. If $\alpha = t_{g,h,i}$, then $\alpha < \beta$ because β is either $t_{g,h,j}$ or $t_{g,h,2n+4-j}$ (if $j \notin \{g,h\}$), or $\beta = t_{g,h,n+2}$ (if $h = j$) or $\beta = t_{g,h,n+1}$ or $t_{g,h,n+3}$ (if $g = j$); similarly, if $\alpha = t_{g,h,2n+4-i}$ then $\alpha > \beta$. Therefore, if $i \notin \{g,h\}$, then there is equal probability that $\alpha < \beta$ and $\alpha > \beta$. Similarly, the same is true if $j \notin \{g,h\}$.

Suppose that $i \in \{g,h\}$ and $j \in \{g,h\}$. Since $i < j$ and $g < h$, we must have $i = g$ and $j = h$. In this case, $\beta = t_{i,j,n+2}$, and $\alpha = t_{i,j,n+1}$ or $\alpha = t_{i,j,n+3}$, hence $\alpha > \beta$ iff $\alpha = t_{i,j,n+3}$.

The probability that $\alpha > \beta$ and α, β are not both in section (i,j) is

$$\frac{1}{2} \cdot \left(1 - \frac{1}{\binom{n}{2}^2}\right).$$

The probability that $\alpha > \beta$ and α, β are both in section (i,j) is

$$\frac{1}{\binom{n}{2}} \cdot \left(\frac{1}{2\binom{n}{2}} + n2\left(q(i,j) - \frac{1}{2}\right)\right) = \frac{1}{2\binom{n}{2}^2} + \left(q(i,j) - \frac{1}{2}\right).$$

Thus, the total probability that $\alpha > \beta$ is exactly $q(i,j)$ as desired.

This concludes the construction and the proof of the theorem.

7. Conclusions

In this paper we studied the complexity of the Bayesian Unit-Demand Item-Pricing problem with a product distribution. We showed that the decision problem is NP-complete even when the distributions have support size 3 and share the same support $\{0, 1, 3\}$ or when they are identical. We also presented a polynomial-time algorithm for distributions of support size 2.

Several interesting open questions remain. Is there a PTAS for general product distributions? Note that our NP-hardness results do not preclude the existence of an FPTAS. Actually, by adapting techniques from Cai and Daskalakis (2011) we can give an FPTAS for the case when the supports of the distributions are integers in a bounded interval. Moreover, we conjecture that the IID case can be solved in polynomial time when the size of the support is constant. For the related problem of finding an optimal lottery-pricing (or an optimal randomized mechanism), such a polynomial-time algorithm was obtained by Daskalakis and Weinberg (2012) for the IID case with a constant support size. The case of general product distributions, however, was shown recently to be intractable in Chen et al. (2015) under standard complexity-theoretic assumptions.

References

Alaei, S., Fu, H., Haghpanah, N., Hartline, J., 2013. The simple economics of approximately optimal auctions. In: Proceedings of the 54th IEEE Annual Symposium on Foundations of Computer Science, pp. 628–637.

Babaioff, M., Immorlica, N., Lucier, B., Weinberg, S.M., 2014. A simple and approximately optimal mechanism for an additive buyer. In: Proceedings of the 55th IEEE Annual Symposium on Foundations of Computer Science, pp. 21–30.

Balcan, M.-F., Blum, A., Hartline, J., Mansour, Y., 2008. Reducing mechanism design to algorithm design via machine learning. J. Comput. Syst. Sci. 74 (8), 1245–1270.

Briest, P., 2008. Uniform budgets and the envy-free pricing problem. In: Proceedings of the 35th International Colloquium on Automata, Languages and Programming, pp. 808–819.

Briest, P., Chawla, S., Kleinberg, R., Weinberg, S.M., 2010. Pricing randomized allocations. In: Proceedings of the 21st Annual ACM–SIAM Symposium on Discrete Algorithms, pp. 585–597.

Cai, Y., Daskalakis, C., 2011. Extreme-value theorems for optimal multidimensional pricing. In: Proceedings of the 52nd Annual IEEE Symposium on Foundations of Computer Science, pp. 522–531.

Cai, Y., Devanur, N.R., Weinberg, S.M., 2016. A duality based unified approach to Bayesian mechanism design. In: Proceedings of the 48th Annual ACM Symposium on Theory of Computing, pp. 926–939.

Cai, Y., Huang, Z., 2013. Simple and nearly optimal multi-item auctions. In: Proceedings of the 24th Annual ACM–SIAM Symposium on Discrete Algorithms, pp. 564–577.

Chawla, S., Hartline, J.D., Kleinberg, R.D., 2007. Algorithmic pricing via virtual valuations. In: Proceedings of the 8th ACM Conference on Electronic Commerce, pp. 243–251.

Chawla, S., Hartline, J.D., Malec, D.L., Sivan, B., 2010. Multi-parameter mechanism design and sequential posted pricing. In: Proceedings of the 42nd Annual ACM Symposium on Theory of Computing, pp. 311–320.

Chawla, S., Malec, D.L., Sivan, B., 2015. The power of randomness in Bayesian optimal mechanism design. Games Econ. Behav. 91, 297–317.

Chawla, S., Miller, J.B., 2016. Mechanism design for subadditive agents via an ex ante relaxation. In: Proceedings of the 2016 ACM Conference on Economics and Computation, pp. 579–596.

Chen, X., Diakonikolas, I., Orfanou, A., Paparas, D., Sun, X., Yannakakis, M., 2015. On the complexity of optimal lottery pricing and randomized mechanisms. In: Proceedings of the 56th IEEE Annual Symposium on Foundations of Computer Science, pp. 1464–1479.

Daskalakis, C., Deckelbaum, A., Tzamos, C., 2012. Optimal pricing is hard. In: Proceedings of the 8th Workshop on Internet and Network Economics, pp. 298–308.

Daskalakis, C., Deckelbaum, A., Tzamos, C., 2014. The complexity of optimal mechanism design. In: Proceedings of the 25th Annual ACM–SIAM Symposium on Discrete Algorithms, pp. 1302–1318.

Daskalakis, C., Weinberg, S.M., 2012. Symmetries and optimal multi-dimensional mechanism design. In: Proceedings of the 13th ACM Conference on Electronic Commerce, pp. 370–387.

Devanur, N.R., Kannan, R., 2008. Market equilibria in polynomial time for fixed number of goods or agents. In: Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science, pp. 45–53.

Etessami, K., Yannakakis, M., 2010. On the complexity of Nash equilibria and other fixed points. SIAM J. Comput. 39 (6), 2531–2597. Preliminary version in Proc. 48th IEEE FOCS, 2007.

Garey, M.R., Johnson, D.S., 1990. Computers and Intractability; A Guide to the Theory of NP-Completeness. W.H. Freeman, New York, NY, USA.

Giannakopoulos, Y., Koutsoupias, E., 2014. Duality and optimality of auctions for uniform distributions. In: Proceedings of the 15th ACM Conference on Economics and Computation, pp. 259–276.

Guruswami, V., Hartline, J.D., Karlin, A.R., Kempe, D., Kenyon, C., McSherry, F., 2005. On profit-maximizing envy-free pricing. In: Proceedings of the 16th Annual ACM–SIAM Symposium on Discrete Algorithms, pp. 1164–1173.

Haghpanah, N., Hartline, J., 2015. Reverse mechanism design. In: Proceedings of the 16th ACM Conference on Electronic Commerce, pp. 757–758.

Hart, S., Nisan, N., 2012. Approximate revenue maximization with multiple items. In: Proceedings of the 13th ACM Conference on Electronic Commerce, p. 656.

Hart, S., Nisan, N., 2013. The menu-size complexity of auctions. In: Proceedings of the 14th ACM Conference on Electronic Commerce, pp. 565–566.

Hartline, J.D., Koltun, V., 2005. Near-optimal pricing in near-linear time. In: Proceedings of 9th International Workshop on Algorithms and Data Structures, pp. 422–431.

Li, X., Yao, A.C.-C., 2013. On revenue maximization for selling multiple independently distributed items. Proc. Natl. Acad. Sci. 110 (28), 11232–11237.

Manelli, A.M., Vincent, D.R., 2007. Multidimensional mechanism design: revenue maximization and the multiple-good monopoly. J. Econ. Theory 137 (1), 153–185.

Myerson, R.B., 1981. Optimal auction design. Math. Oper. Res., 58–73.

Papadimitriou, C.H., 1994. Computational Complexity. Addison-Wesley.

Pavlov, G., 2010. A property of solutions to linear monopoly problems. B.E. J. Theoretical Econ. 11 (1).

Rubinstein, A., Weinberg, S.M., 2015. Simple mechanisms for a subadditive buyer and applications to revenue monotonicity. In: Proceedings of the 16th ACM Conference on Economics and Computation, pp. 377–394.

Tang, P., Wang, Z., 2017. Optimal mechanisms with simple menus. J. Math. Econ. 69, 54–70.

Thanassoulis, J., 2004. Haggling over substitutes. J. Econ. Theory 117 (2), 217–245.

Yao, A.C.-C., 2015. An n-to-1 bidder reduction for multi-item auctions and its applications. In: Proceedings of the 26th Annual ACM–SIAM Symposium on Discrete Algorithms, pp. 92–109.