The Complexity of Optimal Multidimensional Pricing

Xi Chen*   Ilias Diakonikolas†   Dimitris Paparas‡   Xiaorui Sun*

Mihalis Yannakakis‡

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 then 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 $\mathbf{v} = (v_1, \ldots, v_n)$ 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. We assume, as is commonly the case, that the buyer here is *quasi-linear*, i.e., her utility for item $i \in [n]$ is $v_i - p_i$, and she will select an item with the maximum nonnegative utility or nothing if no such item exists. This is known as the *Bayesian Unit-demand Item-Pricing Problem* (*BUPP*) [CHK07], and has received considerable attention in the CS literature during the past few years [GHK+05, CHK07, Bri08, CHMS10, CD11, DDT14].

Throughout this paper we focus on the well-studied case [CHK07, CHMS10, CD11] that $\mathcal{F} = \times_{i=1}^{n} \mathcal{F}_i$ is a *product distribution*, i.e., the valuations of the buyer for the n items are mutually independent random variables. We assume that the n (marginal) distributions $\mathcal{F}_i$ are discrete and are known to the seller, i.e., the values of the support and the corresponding probabilities are rational numbers given explicitly in the input.

This seemingly simple computational problem appears to exhibit a very rich structure. Prior to our work, even the (very special) case that the distributions $\mathcal{F}_i$ have support 2 was not well understood: First note that the search space is apparently exponential, since the support size of $\mathcal{F}$ is 2^n. What makes things trickier is that the optimal prices are not necessarily in the support of $\mathcal{F}$ (see [CD11] for a simple example with two items with distributions of support 2). So, a priori, it was not even clear whether the optimal prices can be described with polynomially many bits in the size of the input description.

Revenue-optimal pricing is well-studied by economists (see, e.g., [Wil96] for a survey and [MMW89] for a simple additive case with two items). The pricing problem studied in this work fits in the general framework of *optimal multi-dimensional mechanism design*, a central question in mathematical economics (see [MV07] and references therein). Finding the optimal deterministic mechanism in our setting is equivalent to finding the optimal item-pricing. A randomized mechanism, on the other hand, would allow the seller to price lotteries over items [BCKW10, CMS10], albeit this may be less natural in this context.

Optimal mechanism design is well-understood in single-parameter settings for which Myerson [Mye81] gives a closed-form characterization for the optimal mechanism. Chawla, Hartline and Kleinberg [CHK07] show that techniques from Myerson's work can be used to obtain an analogous closed-form characterization (and also an efficient algorithm) for pricing in our setting, albeit with a constant factor loss in the revenue. In particular, they obtain a factor 3 approximation to the optimal expected revenue (subsequently improved to 2 in [CHMS10]). Cai and Daskalakis [CD11] obtain a polynomial-time approximation scheme for distributions with monotone hazard-rate (and a quasi-polynomial time approximation scheme for the broader

class of regular distributions). That is, prior to this work, closed-form characterizations (and efficient algorithms) were known for *approximately* optimal pricing. The question of whether such a characterization exists for the *optimal* pricing has remained open and was posed as an open problem in these works [CHK07, CD11].

Our Results. In this paper, we take a principled complexity-theoretic look at the BUPP with independent (discrete) distributions. We start by showing (Theorem 1) that the general decision problem is in NP (and as a corollary, the optimal prices can be described with polynomially many bits). We note that the membership proof is non-trivial because the optimal prices may not be in the support. Our proof proceeds by partitioning the space of price-vectors into a set of (exponentially many) cells (defined by the value distributions $\mathcal{F}_i$), so that the optimal revenue within each cell can be found efficiently by a shortest path computation. One consequence of the analysis is that the optimal pricing problem has the integrality property: if the values in the supports are integer then the optimal prices are also integer (though they may not belong to the support).

We then proceed to show (Theorem 2) that the case in which each marginal distribution has support at most 2 can be solved in polynomial time. Indeed, by exploiting the underlying structure of the problem, we show that it suffices to consider $O(n^2)$ price-vectors to compute the optimal revenue in this case.

Our main result is that the problem is NP-hard, even for distributions of support 3 (Theorem 3) *or distributions that are identical but have large support* (Theorem 4). This answers an open problem first posed in [CHK07] and also asked in [CD11, DDT14]. The main difficulty in the reductions stems from the fact that, for a general instance of the pricing problem, the expected revenue is a highly complex nonlinear function of the prices. The challenge is to construct an instance such that the revenue can be well-approximated by a simple function and is also general enough to encode an NP-hard problem.

Previous Work. We have already mentioned the main algorithmic works for the independent distributions case with approximately-optimal revenue guarantees [CHK07, CHMS10, CD11]. On the lower bound side, Guruswami et al. [GHK+05] and subsequently Briest [Bri08] studied the complexity of the problem when the buyer's values for the items are *correlated*, respectively obtaining APX-hardness and $\Omega(n^\epsilon)$ inapproximability, for some constant $\epsilon > 0$. More recently, Daskalakis, Deckelbaum and Tzamos [DDT14] showed that the pricing problem with independent distributions is SQRT-SUM-hard 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, it is SQRT-SUM-hard to compare the revenue of two specific price-vectors. This has no bearing on the complexity of the problem under the standard discrete model we consider here, since the exact revenue of a price-vector can be computed efficiently.

Related Work. The optimal mechanism design problem, i.e., the problem of finding a revenue-maximizing mechanism in a Bayesian setting, has received considerable attention in the CS community during the past few years. The vast majority of the work so far is algorithmic [CHK07, CHMS10, BGGM10, Ala11, DFK11, HN12, CDW12a, CDW12b], providing approximation or exact algorithms for various versions of the problem. Regarding lower bounds, Papadimitriou and Pierrakos [PP11] show that computing the optimal deterministic single-item auction is APX-hard, even for the case of 3 bidders. We remark that, if randomization is allowed, then this problem can be solved exactly in polynomial time via linear programming [DFK11]. In a very recent work [DDT12], Daskalakis, Deckelbaum and Tzamos show #P-hardness of computing the optimal randomized mechanism for the case of additive buyers. We remark that their result does not have any implication for the unit-demand case due to the very different structures of the two problems.

The rest of the paper is organized as follows. In Section 2 we define formally the problem, state our main results, and prove some preliminary basic properties. In Section 3 we show that the decision problem is in NP. In Section 4 we present 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. We conclude in Section 7.

2 Preliminaries

In our setting, there are one buyer and one seller with n items, indexed by $[n] = \{1, 2, \ldots, n\}$. The buyer is interested in buying at most one item (unit demand), and her valuation of the items are drawn from n independent discrete distributions, one for each item. We use $V_i = \{v_{i,1}, \ldots, v_{i,|V_i|}\}$, for each $i \in [n]$, to denote the support of the value distribution of item i, where $0 \leq v_{i,1} < \cdots < v_{i,|V_i|}$. 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 also use $\Pr[\mathbf{v}]$ to denote the probability of the valuation vector being $\mathbf{v} = (v_1, \ldots, v_n) \in V$, 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 $q_{i,j}$'s, are given to the seller explicitly. The seller then assigns a non-negative price p_i to each item $i \in [n]$. Once the price vector $\mathbf{p} = (p_1, \ldots, p_n)$ is determined, the buyer draws her values $\mathbf{v} = (v_1, \ldots, v_n)$ from the distributions independently, i.e., her values are $\mathbf{v} \in V$ with probability $\Pr[\mathbf{v}]$. We assume that the buyer is quasi-linear: her utility for item i equals $v_i - p_i$. Let

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

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. For convenience let $\mathcal{T}(\mathbf{v}, \mathbf{p})$ denote the set of items with maximum non-negative utility ($\mathcal{T}(\mathbf{v}, \mathbf{p}) = \emptyset$ iff $\mathcal{U}(\mathbf{v}, \mathbf{p}) < 0$).

Knowing the distributions (as well as the behavior of the buyer), the seller's objective is then to compute a vector $\mathbf{p} \in \mathbb{R}_+^n$ that maximizes her 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 to denote the following decision problem: The input consists of n discrete distributions, where $v_{i,j}$ and $q_{i,j}$ are all rational and encoded in binary, and a rational number $t \geq 0$. The problem asks whether the supremum of the expected revenue $\mathcal{R}(\mathbf{p})$ over all price vectors $\mathbf{p} \in \mathbb{R}_+^n$ is at least t, where we use $\mathbb{R}_+$ to denote the set of non-negative real numbers.

We note that the aforementioned decision problem is 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 non-negative utility. Throughout the paper, we will use the following *maximum price*[1] tie-breaking rule (which is convenient for our arguments): when there are multiple items with maximum non-negative 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 so that it makes sense to talk about "the" item selected by the buyer in the proofs.) We will use $\mathcal{R}(\mathbf{v}, \mathbf{p})$ to denote the seller's revenue under the maximum price tie-breaking rule when the valuation vector is $\mathbf{v} \in V$. So we have

$$\mathcal{R}(\mathbf{p}) = \sum_{\mathbf{v} \in V} \Pr[\mathbf{v}] \cdot \mathcal{R}(\mathbf{v}, \mathbf{p}).$$

---

[1] 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 show in Section 2.2 of the full version that our choice of the maximum price tie-breaking rule does not affect the supremum of the expected revenue (hence, the complexity of the problem):

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

We will henceforth always adopt the maximum price tie-breaking rule throughout the rest of the paper, and let $\mathcal{R}(\mathbf{v}, \mathbf{p})$ denote the revenue of the seller with respect to this rule. We show two more lemmas in Section 2.2 of the full paper. Given V_i, let $P = \times_{i=1}^n [a_i, b_i]$, where $a_i = \min_j v_{i,j}$ and $b_i = \max_j v_{i,j}$.

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

LEMMA 2.3. *There exists a vector $\mathbf{p}^* \in P$ such that*

$$\mathcal{R}(\mathbf{p}^*) = \sup_{\mathbf{p} \in \mathbb{R}_+^n} \mathcal{R}(\mathbf{p}).$$

By Lemma 2.3, one of the advantages of the maximum price rule is that the supremum of the expected revenue $\mathcal{R}(\mathbf{p})$ is always achievable, so it makes sense to talk about whether a price vector is optimal or not. In the following example, we point out that this does not hold for general tie-breaking rules.

Example: Suppose item 1 has value 10 with probability 1, item 2 has value 8 with probability $1/2$ and value 12 with probability $1/2$, and in case of tie the buyer prefers item 1. The supremum in this example is 11: set $p_1 = 10$ for item 1 and $p_2 = 12 - \epsilon$ for item 2. The buyer will buy item 1 with probability $1/2$ (if her value for item 2 is 8) and item 2 with probability $1/2$ (if her value for item 2 is 12). However, an expected revenue of 11 is not achievable: if we give price 12 to item 2, then the buyer will always buy item 1 and the revenue is 10. Note that the expected revenue for this tie-breaking rule is not a continuous function of the prices.

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

THEOREM 1. ITEM-PRICING *is in NP.*

Second, we present in Section 4 a polynomial-time algorithm for ITEM-PRICING when all the distributions have support size at most 2.

THEOREM 2. ITEM-PRICING *is in P when every distribution has support size at most 2.*

As our main result, we resolve the complexity of the problem, by showing that it is NP-hard even when all distributions have support at most 3 (Section 5), or when they are identical (Section 6).

THEOREM 3. ITEM-PRICING *is NP-hard even when every distribution has support size at most* 3.

THEOREM 4. ITEM-PRICING *is NP-hard even when the distributions are identical.*

3 Membership in NP

In this section we prove Theorem 1, i.e., ITEM-PRICING is in NP.

Proof. [Proof Sketch of Theorem 1] We will partition $P = \times_{i=1}^{n}[a_i, b_i]$ into equivalence classes, so that two price vectors $\mathbf{p}, \mathbf{p}'$ from the same class yield the same outcome for all valuations $\mathbf{v}$, i.e., the buyer selects the same item or none at all. Consider the partition of P induced by the following set of hyperplanes. For each item i 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). We can show then:

CLAIM 3.1. *For every valuation $\mathbf{v} \in V$, all the price vectors in each cell yield the same outcome.*

Next we show that, for each cell C, it is easy to compute the supremum of the expected revenue $\mathcal{R}(\mathbf{p})$ over $\mathbf{p} \in C$. To this end, we let W_i denote the set of valuations for which the buyer picks item i if the prices lie in the cell C, and let γ_i be the probability of W_i: $\gamma_i = \sum_{\mathbf{v} \in W_i} \Pr[\mathbf{v}]$. By Claim 3.1, W_i and γ_i are the same for all prices in the cell C. It turns out that γ_i can be computed efficiently as follows. For each $s_i \in V_i$, let $V(s_i)$ be the set of valuations with $v_i = s_i$ for which the buyer picks item i if the prices lie in the cell C. Then W_i is the disjoint union of $V(s_i)$, $s_i \in V_i$. For each $j \neq i$, we can determine efficiently the subset of values $L_j \subseteq V_j$ such that the buyer prefers item i to j if i has value s_i and j has value from L_j. $V(s_i)$ is the Cartesian product of L_j, $j \neq i$, and $\{s_i\}$. Thus, we multiply the probabilities of L_j's and the probability of s_i. Summing up the probabilities of $V(s_i)$ over $s_i \in V_i$ gives us γ_i.

Finally, the supremum of the expected revenue $\mathcal{R}(\mathbf{p})$ over $\mathbf{p} \in C$ is the maximum of $\sum_{i \in [n]} \gamma_i \cdot p_i$, 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 $\sum_i \gamma_i \cdot p_i$ over all $\mathbf{p} \in C$ can be computed in polynomial time by solving the linear program that maximizes $\sum_i \gamma_i \cdot p_i$ subject to $\mathbf{p} \in C'$. In fact, this LP has a special form and as we will show below, we can test

feasibility by solving a negative weight cycle problem, and we can compute the optimal solution by solving a single-source shortest path problem. It follows that the specification of a cell C in the partition is an appropriate *yes* certificate for the decision problem ITEM-PRICING, and the theorem is proved.

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 $p_i \leq s_i$, 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 $p_i \geq s_i$ (or only for the tightest one) include an edge $(i, 0)$ with weight $-s_i$. For each inequality of the form $p_i - p_j < s_i - t_j$ or $p_i - p_j \leq s_i - t_j$ (or only for the tightest), include a (strict or weak) edge (j, i) with weight $s_i - t_j$. Similarly, for each inequality of the form $p_i - p_j > s_i - t_j$ or $p_i - p_j \geq s_i - t_j$ (or only for the tightest), include a (strict or weak) edge (i, j) with weight $t_j - s_i$. We have the following connections:

LEMMA 3.1. *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 achieved by the price vector $\mathbf{p}$ that consists of the distances from node 0 to the other nodes of the graph G.

The NP characterization of ITEM-PRICING as well as the corresponding structural characterization of the optimal price vector of each cell have several easy and useful consequences. First, we get an alternative proof of Lemma 2.3 (see the full version). Another consequence suggested by the characterization of Lemma 3.1 is that the maximum expected revenue can always be achieved by a price vector $\mathbf{p}$ in which all p_i are sums of a value and differences between pairs of values of items. This implies the following useful corollary.

COROLLARY 3.1. *If all the values in V_i, $i \in [n]$, are integers, then there must exist an optimal price vector $\mathbf{p} \in P$ with integer coordinates.*

4 A Polynomial-Time Algorithm for Support 2

In this section we present a polynomial-time algorithm for the case that each distribution has 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 generalize this algorithm to handle the general case.

4.1 An Interesting Special Case In this subsection, we assume that every item i has support size exactly 2, where $V_i = \{a_i, b_i\}$ satisfies $b_i > a_i > 0$, for all i. Let $q_i : 0 < q_i < 1$ denote the probability of the value of item i being b_i, and let $t_i = b_i - a_i > 0$. Moreover, we assume in this subsection that the two vectors $\mathbf{a} = (a_1, \ldots, a_n)$ and $\mathbf{b} = (b_1, \ldots, b_n)$ satisfy the following **"non-degeneracy" assumption**: $b_1 < \cdots < b_n$, $a_i \neq a_j$ 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. We prove a sequence of lemmas below (all the proofs, except that of Lemma 4.5, can be found in the full version) to show that, given $\mathbf{a}$ and $\mathbf{b}$ that satisfy all the conditions above, one can compute efficiently a set $A \subset P$ of price vectors (independent of q_i's) such that $|A| = O(n^2)$ and OPT $\subseteq A$. As a result, by computing $\mathcal{R}(\mathbf{p})$ for all $\mathbf{p} \in A$ we obtain the maximum of expected revenue and an optimal price vector.

We start with the following two lemmas:

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 $\mathbf{p} \notin$ OPT.*

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

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 $i \in [n]$.

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 $\mathbf{p} \notin$ OPT.*

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 $t_i < t_k$ ($t_i \neq t_k$ from 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 $\mathbf{p} \notin$ OPT.*

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 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. 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 $A \subset P$ to denote this set of 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 $\mathbf{p} \notin$ 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. 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 $\mathbf{p} \notin$ OPT.

We need some notation. Let V' denote the projection of V onto all but the cth and dth coordinates, i.e., $V' = \times_{i \in [n] - \{c,d\}} V_i$. We still use $[n] - \{c, d\}$ to index entries of vectors $\mathbf{u}$ in V'. Let $U \subseteq V'$ denote the set of $\mathbf{u} \in V'$ such that $u_i - p_i < t$ for all $i > d$. (This simply 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 an 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 $\mathbf{v} = (\mathbf{u}, v_c, v_d) \in V$ satisfies $\mathbf{u} \notin U$, 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 $\mathbf{v} = (\mathbf{u}, a_c, a_d) \in V$, $\mathcal{R}(\mathbf{v}, \mathbf{p}) = \mathcal{R}(\mathbf{v}, \mathbf{p}') = \mathcal{R}(\mathbf{v}, \mathbf{p}^*)$ as the utility from c and d are negative.

Now we consider a price 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, given any $\mathbf{u} \in U$, 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 $(\mathbf{u}, b_c, b_d)$. From the definition of U, we have the following simple equations: For $\mathbf{p}$, we have

$$\mathcal{R}(\mathbf{u}_2^+, \mathbf{p}) = b_c - t \quad \text{and} \quad \mathcal{R}(\mathbf{u}_3^+, \mathbf{p}) = b_c - t;$$

For $\mathbf{p}'$, we have $\mathcal{R}(\mathbf{u}_2^+, \mathbf{p}') = b_c - t$ and

$$\mathcal{R}(\mathbf{u}_1^+, \mathbf{p}') = \mathcal{R}(\mathbf{u}_3^+, \mathbf{p}') = b_d - t.$$

The following properties are also easy to verify:

$$(4.1) \qquad \mathcal{R}(\mathbf{u}_1^+, \mathbf{p}) = \mathcal{R}(\mathbf{u}_1^+, \mathbf{p}^*) = \mathcal{R}(\mathbf{u}_3^+, \mathbf{p}^*)$$
$$\mathcal{R}(\mathbf{u}_1^+, \mathbf{p}^*) - (b_d - b_c) \leq \mathcal{R}(\mathbf{u}_2^+, \mathbf{p}^*) \leq \mathcal{R}(\mathbf{u}_1^+, \mathbf{p}^*).$$

Recall that we use $\Pr[\mathbf{v}]$ to denote the probability of the valuation vector being $\mathbf{v} \in V$. 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

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

Then we have $h_1, h_2, h_3 > 0$ and $\Pr[\mathbf{u}_i^+] = \Pr[\mathbf{u}] \cdot h_i$.

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

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

We can rewrite the first two sums for $\mathbf{p}$ and $\mathbf{p}'$ as follows (here all sums are over $\mathbf{u} \in U$):

$$h_1 \sum_{\mathbf{u}} \Pr[\mathbf{u}] \cdot \mathcal{R}(\mathbf{u}_1^+, \mathbf{p}) + h_2 \sum_{\mathbf{u}} \Pr[\mathbf{u}] \cdot (b_c - t)$$
$$(4.2) \qquad\qquad + h_3 \sum_{\mathbf{u}} \Pr[\mathbf{u}] \cdot (b_c - t),$$

$$h_1 \sum_{\mathbf{u}} \Pr[\mathbf{u}] \cdot (b_d - t) + h_2 \sum_{\mathbf{u}} \Pr[\mathbf{u}] \cdot (b_c - t)$$
$$(4.3) \qquad\qquad + h_3 \sum_{\mathbf{u}} \Pr[\mathbf{u}] \cdot (b_d - t).$$

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

$$(4.4) \qquad \sum_{\mathbf{u}} \Pr[\mathbf{u}] \cdot \mathcal{R}(\mathbf{u}_1^+, \mathbf{p}) > \sum_{\mathbf{u}} \Pr[\mathbf{u}] \cdot (b_d - t).$$

We can also rewrite the sum for $\mathcal{R}(\mathbf{p}^*)$ as

$$h_1 \sum_{\mathbf{u}} \Pr[\mathbf{u}] \cdot \mathcal{R}(\mathbf{u}_1^+, \mathbf{p}^*) + h_2 \sum_{\mathbf{u}} \Pr[\mathbf{u}] \cdot \mathcal{R}(\mathbf{u}_2^+, \mathbf{p}^*)$$
$$(4.5) \qquad\qquad + h_3 \sum_{\mathbf{u}} \Pr[\mathbf{u}] \cdot \mathcal{R}(\mathbf{u}_3^+, \mathbf{p}^*).$$

The first sum in (4.5) is the same as that of (4.2). By (4.1) and (4.4), we have

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

The third sum in (4.5) is also strictly larger than that of (4.2) as $\mathcal{R}(\mathbf{u}_3^+, \mathbf{p}^*) = \mathcal{R}(\mathbf{u}_1^+, \mathbf{p}^*) \geq \mathcal{R}(\mathbf{u}_2^+, \mathbf{p}^*)$, while the second and third sums in (4.2) are the same, ignoring h_2 and h_3. As a result, we have $\mathcal{R}(\mathbf{p}^*) > \mathcal{R}(\mathbf{p})$.

4.2 General Case In this subsection, we deal with the general case. Let I denote an input instance with n items, in which $|V_i| \leq 2$ for all i. For each $i \in [n]$ either $V_i = \{a_i, b_i\}$ with $b_i > a_i \geq 0$, or $V_i = \{b_i\}$ with $b_i \geq 0$.

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.) Since permuting the items does not affect the maximum expected revenue, we assume without loss of generality that $b_1 \leq \cdots \leq b_n$.

The idea is to perturb the instance I (symbolically) so that the new instances satisfy all the 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 let $\mathcal{R}(\mathbf{p})$ and $\mathcal{R}_\epsilon(\mathbf{p})$ denote the revenue with respect to I and I_ϵ, respectively.

It is easy to verify that, when ϵ is sufficiently small, the new instance I_ϵ satisfies all conditions given at the beginning of Section 4.1, including the non-degeneracy assumption. Moreover, we show that

LEMMA 4.6. *The limit of* $\max_{\mathbf{p}} \mathcal{R}_\epsilon(\mathbf{p})$ *exists as* $\epsilon \to 0$, *and can be computed in polynomial time.*

Proof. Because I_ϵ satisfies all the conditions, we know there is a set of $O(n^2)$ price vectors, denoted by A_ϵ for I_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 $\mathbf{p}_\epsilon$ in A_ϵ has an explicit expression in ϵ: each entry of $\mathbf{p}_\epsilon$ is indeed an affine linear function of ϵ. As a result, the limit of $\mathcal{R}_\epsilon(\mathbf{p}_\epsilon)$ as ϵ approaches 0 exists and

can be computed efficiently. Since $\lim_{\epsilon \to 0}(\max_{\mathbf{p}} \mathcal{R}_\epsilon(\mathbf{p}))$ is just the maximum of these $O(n^2)$ limits, it also exists and can be computed in polynomial time in the input size of I.

The next lemma shows that this limit is exactly the maximum expected revenue of I. The proof can be found in the full version.

LEMMA 4.7. $\max_{\mathbf{p}} \mathcal{R}(\mathbf{p}) = \lim_{\epsilon \to 0} \left(\max_{\mathbf{p}} \mathcal{R}_\epsilon(\mathbf{p}) \right)$.

5 NP–Hardness for Support Size 3

In this section we present a polynomial-time reduction from PARTITION to ITEM-PRICING when each distribution has support (at most) 3. Recall that in the PARTITION problem [GJ79] we are given a set $C = \{c_1, \ldots, c_n\}$ of n positive integers and wish to determine whether it is possible to partition C into two subsets with equal sum. We may assume without loss of generality that $c_1 = \max(c_1, \ldots, c_n)$.

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

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

We will eventually set $a = 1$ and $b = 3$, but for the sake of the presentation, we keep a, b as generic constants till the end. Note that the definition of r_i implies that

$$(5.6) \qquad bq_i = a(q_i + r_i) - ar_i t_i.$$

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 $\text{poly}(n)/N$, 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 C 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.

Before we proceed with the proof, we will need some notation. For $T_1, T_2, \epsilon \in \mathbb{R}_+$ we write $T_1 = T_2 \pm \epsilon$ to denote 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 have $r_i = (b - a)q_i/a \pm O(n/N^2)$.

Lemma 2.2 and Corollary 3.1 imply that a revenue maximizing price vector can be assumed to have non-negative integer coordinates 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 in* $\{a, b\}^n$.

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

Approximating the expected revenue. We appropriately partition V into three disjoint 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 three disjoint events:

- *First Event*: $E_1 = \{\mathbf{v} \in V \mid \exists i \in S : v_i = b\}$. Note that $\mathcal{R}(\mathbf{v}, \mathbf{p}) = a$ for any $\mathbf{v} \in E_1$. We have

$$(5.7) \qquad \Pr[E_1] = 1 - \prod_{i \in S}(1 - q_i)$$

$$= \sum_{i \in S} q_i - \sum_{i \neq j \in S} q_i q_j \pm O(\epsilon).$$

- *Second Event*: $E_2 = \overline{E_1} \cap \{\mathbf{v} \in V \mid \exists i \in S : v_i = a \text{ and } \forall i \in T : v_i \in \{0, a\}\}$. For any $\mathbf{v} \in E_2$ we have $\mathcal{R}(\mathbf{v}, \mathbf{p}) = a$. The probability $\Pr[E_2]$ is

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

$$= \sum_{i \in S} r_i - \sum_{i \in S} r_i \sum_{j \in T} q_j + \sum_{i \neq j \in S} q_i q_j$$

$$- \sum_{i \neq j \in S}(q_i + r_i)(q_j + r_j) \pm O(\epsilon).$$

- *Third Event:* $E_3 = \overline{E_1} \cap \{\mathbf{v} \in V \mid \exists i \in T : v_i = b\}$. We have $\mathcal{R}(\mathbf{v}, \mathbf{p}) = b$ for any valuation $\mathbf{v} \in E_3$, and the probability $\Pr[E_3]$ of this event is

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

The expected revenue is then

$$\mathcal{R}(\mathbf{p}) = \big(\Pr[E_1] + \Pr[E_2]\big) \cdot a + \Pr[E_3] \cdot b.$$

With (5.6), after performing a number of rearrangements of the terms and simplifications, and setting

$$L = b \sum_{j \in [n]} q_j - b \sum_{i \neq j \in [n]} q_i q_j$$

(notice that L does not depend on the pricing, i.e., the partition of the items into S and T), we eventually have that $\mathcal{R}(\mathbf{p})$ is equal to:

$$L + \sum_{i \in S} r_i \left(a t_i - \frac{b}{2} \sum_{j \in S, j \neq i} q_j - a \sum_{j \in T} q_j \right) \pm O(\epsilon)$$
$$= L + \frac{1}{M^2} \left(\sum_{i \in S} c_i \right) \left(\sum_{j \in T} c_j \right) \pm O(\epsilon),$$

by setting $a = 1, b = 3$ in the last expression. Details of the calculations can be found in the full paper.

At this point, we observe that the sum of the two factors $\sum_{i \in S} c_i$ and $\sum_{j \in T} c_j$ in the equation above 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 these two factors is maximized. In particular, if there exists a partition of $C = \{c_1, \ldots, c_n\}$ 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 the set C, then for any partition, the revenue will be at most

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

As $\epsilon = o(1/M^2)$ 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

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

This completes the proof sketch.

Remark. In the construction above, the support of the distributions includes the value 0. It is easy to modify the construction, if desired, so that the support contains only positive values: Shift all values of the distributions up by 1 (thus, the supports now become $\{1, 2, 4\}$) 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 that ITEM-PRICING is NP-hard even for identical distributions. For this purpose, we reduce from the following (still NP-complete) version of Integer Knapsack.

DEFINITION 6.1. (INT. KNAPSACK WITH REPETITIONS) INPUT: $n + 1$ *positive integers* $a_1 < \cdots < a_n$ *and* L. PROBLEM: *Do there exist nonnegative integers* $x_1, \ldots, x_n$ *s.t.* $\sum_{i \in [n]} x_i = n$ *and* $\sum_{i \in [n]} x_i a_i = L$?

Given an instance of this Knapsack problem, we reduce it to an instance of ITEM-PRICING with n items, each of which has its value drawn independently from a suitably constructed distribution Q over nonnegative integers. Similar to the reduction for support size 3, Q assigns most of its probability mass to the point 0, so that valuations with many nonzero values contribute very little to the expected revenue. We set the support and probabilities of Q carefully, so that the optimization of the expected revenue amounts to a quadratic optimization problem that mimics the Integer Knapsack problem with repetitions. The construction and the proof are quite involved (see the full paper), so we will only give an outline here.

Let $m = \max(n^5, a_n)$ and $N = m^{n^2}$. For each $i \in [n]$, let $v_i = m^{n+i}$. For each $i \in [n-1]$, let

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

Let $\gamma_n = 1/(Nm^{2n})$. Let $\Gamma_i = \sum_{j=i}^n \gamma_j = 1/(Nm^{n+i})$, for each i. Note that $v_i \Gamma_i = 1/N$ for all i.

The construction of Q also uses a sequence of probability distributions $\mathbf{q}_1, \ldots, \mathbf{q}_n$, whose supports are all subsets of $[2n^3]$, as well as a sequence of (not necessarily positive) numbers $t_1, \ldots, t_n$ with $|t_i| = O(1/N^2)$ for all $i \in [n]$. We delay the definition of $\mathbf{q}_i$ and t_i for now and define Q using v_i, γ_i, t_i and $\mathbf{q}_i$.

First the support of Q is

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

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, let $T_i = \sum_{j=i}^n t_j$, and

$$r_i = \sum_{j=i}^n (\gamma_j + t_j) = \Gamma_i + T_i, \quad \text{for each } i \in [n].$$

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, as long as $|t_i| = O(1/N^2)$ for each $i \in [n]$, we can prove the following useful lemma about optimal price vectors.

LEMMA 6.1. *There is an optimal price vector in* $\{v_1, \ldots, v_n\}^n$.

Given $\mathbf{p} \in \{v_1, \ldots, v_n\}^n$, let x_i denote the number of items priced at v_i. Then $\sum_i x_i = n$. We will only consider the contribution to $\mathcal{R}(\mathbf{p})$ of valuation vectors with at most two positive entries. The following lemma shows that the contribution from other valuations is of third order with respect to (roughly) $1/N$.

LEMMA 6.2. *The revenue from valuation vectors with at least three positive entries is* $O(n^3/(m^{n+3}N^3))$.

Let $\epsilon' = n^3 m^{n-1}/N^3$. Then a careful analysis of the contribution from valuations with at most two positive entries yields the following approximation of $\mathcal{R}(\mathbf{p})$ after a number of simplifications, with an error of $O(\epsilon')$:

$$(6.8) \quad \frac{n}{N} + \sum_{i \in [n]} x_i v_i T_i - \frac{n-1}{2N} \sum_{i \in [n]} x_i r_i$$
$$+ \sum_{i < j \in [n]} \frac{x_i x_j}{N} \cdot \left(\frac{1}{2} - p(i,j)\right)(\Gamma_i - \Gamma_j),$$

where, for each pair $i < j \in [n]$, we use $p(i,j)$ to denote the probability that $\alpha - v_i > \beta - v_j$ where α and β are drawn independently from Q, conditioning on $\alpha \geq v_i$ and $\beta \geq v_j$.

Our ultimate goal is to set t_i's and $\mathbf{q}_i$'s appropriately so that (6.8) by the end has the following form:

$$(6.9) \quad \frac{n}{N} + \frac{L^2}{N^2 m^{3n}} - \frac{1}{N^2 m^{3n}} \left(\sum_{i \in [n]} x_i a_i - L\right)^2.$$

If this is the case, we finally obtain a polynomial-time reduction from the special Knapsack problem to ITEM

PRICING, since the difference between (6.9) and $\mathcal{R}(\mathbf{p})$ is at most $O(\epsilon')$ and thus (6.9) 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 the special Knapsack problem with repetitions.

Comparing (6.9) and (6.8) and using $\sum_{i \in [n]} x_i = n$, we can deduce that our goal is achieved if

$$(6.10) \quad T_i = \frac{1}{v_i} \cdot \left(\frac{(n-1)r_i}{2N} - \frac{na_i^2 - 2a_i L}{N^2 m^{3n}}\right),$$

for all $i \in [n]$ (note that the absolute value of the right side of (6.10) is $O(n/(m^{2n+2}N^2))$), and

$$(6.11) \quad \frac{((1/2) - p(i,j))(\Gamma_i - \Gamma_j)}{N} = \frac{(a_i - a_j)^2}{N^2 m^{3n}},$$

for all pairs $i < j \in [n]$.

For the first condition, we note that the equations (6.10), for all $i \in [n]$, indeed 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 (6.10) holds for all $i \in [n]$. Moreover, it is easy to show that the t_i's are $O(1/N^2)$ as we promised earlier.

The second condition of (6.11) is more difficult to satisfy. It specifies a desired value for every $p(i,j)$. It is easy to see from (6.11) that all the desired values are very close to $1/2$, namely $0 < 1/2 - p(i,j) = o(1/m)$, for all $i < j \in [n]$. We let $q(i,j)$ denote the probability of $\alpha > \beta$, where α is drawn from $\mathbf{q}_i$ and β is drawn from $\mathbf{q}_j$ independently. Then one can show that there is a linear relation between $p(i,j)$ and $q(i,j)$, for every $i < j \in [n]$; furthermore, $1/2 - p(i,j) = o(1/m)$ implies that $|1/2 - q(i,j)| = O(1/m)$.

As a result, to obtain the desired values of $p(i,j)$, it suffices to construct the distributions $\mathbf{q}_i$, $i \in [n]$, so that $q(i,j)$'s have the corresponding desired values. For arbitrary $q(i,j)$ this is not always possible, 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 all close to $1/2$: $|q(i,j) - 1/2| = O(1/m)$. We show (constructively) in this case that the desired distributions $\mathbf{q}_i$ exist. The construction is nontrivial and is given in the full paper. This concludes the sketch of the proof.

7 Conclusions

In this paper we studied the complexity of the Bayesian Unit-Demand Item-Pricing problem with independent distributions. We showed that the decision problem is NP-complete when the distributions are of support size 3 or when they are identical. We presented a polynomial time algorithm for distributions of support size 2.

Several interesting open questions remain. Is there a PTAS for general distributions? Note that our NP-hardness results here do not preclude the existence of an FPTAS. Actually, by adapting techniques from [CD11] 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. A related question concerns the complexity of the randomized case (i.e., lottery pricing). We conjecture that this problem is intractable, but new ideas are needed to prove this.

References

[Ala11]   Saeed Alaei. Bayesian combinatorial auctions: Expanding single buyer mechanisms to many buyers. In *Proceedings of the 52nd Annual IEEE Symposium on Foundations of Computer Science*, pages 512–521, 2011.

[BCKW10]   Patrick Briest, Shuchi Chawla, Robert Kleinberg, and S. Matthew Weinberg. Pricing randomized allocations. In *Proceedings of the 21st Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 585–597, 2010.

[BGGM10]   Sayan Bhattacharya, Gagan Goel, Sreenivas Gollapudi, and Kamesh Munagala. Budget constrained auctions with heterogeneous items. In *Proceedings of the 42nd Annual ACM Symposium on Theory of Computing*, pages 379–388, 2010.

[Bri08]   Patrick Briest. Uniform budgets and the envy-free pricing problem. In *Proceedings of the 35th International Colloquium on Automata, Languages and Programming*, pages 808–819, 2008.

[CD11]   Yang Cai and Constantinos Daskalakis. Extreme-value theorems for optimal multidimensional pricing. In *Proceedings of the 52nd Annual IEEE Symposium on Foundations of Computer Science*, pages 522–531, 2011.

[CDW12a]   Yang Cai, Constantinos Daskalakis, and S. Matthew Weinberg. An algorithmic characterization of multi-dimensional mechanisms. In *Proceedings of the 44th Annual ACM Symposium on Theory of Computing*, pages 459–478, 2012.

[CDW12b]   Yang Cai, Constantinos Daskalakis, and S. Matthew Weinberg. Optimal multi-dimensional mechanism design: Reducing revenue to welfare maximization. In *Proceedings of the 53rd Annual IEEE Symposium on Foundations of Computer Science*, pages 130–139, 2012.

[CHK07]   Shuchi Chawla, Jason D. Hartline, and Robert D. Kleinberg. Algorithmic pricing via virtual valuations. In *Proceedings of the 8th ACM Conference on Electronic Commerce*, pages 243–251, 2007.

[CHMS10]   Shuchi Chawla, Jason D. Hartline, David L. Malec, and Balasubramanian Sivan. Multi-parameter mechanism design and sequential posted pricing. In *Proceedings of the 42nd Annual ACM Symposium on Theory of Computing*, pages 311–320, 2010.

[CMS10]   Shuchi Chawla, David L. Malec, and Balasubramanian Sivan. The power of randomness in Bayesian optimal mechanism design. In *Proceedings of the 11th ACM Conference on Electronic Commerce*, pages 149–158, 2010.

[DDT12]   Constantinos Daskalakis, Alan Deckelbaum, and Christos Tzamos. Optimal pricing is hard. In *Proceedings of the 8th Workshop on Internet and Network Economics*, pages 298–308, 2012.

[DDT14]   Constantinos Daskalakis, Alan Deckelbaum, and Christos Tzamos. The complexity of optimal mechanism design. In *Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms*, 2014.

[DFK11]   Shahar Dobzinski, Hu Fu, and Robert D. Kleinberg. Optimal auctions with correlated bidders are easy. In *Proceedings of the 43rd Annual ACM Symposium on Theory of Computing*, pages 129–138, 2011.

[GHK+05]   Venkatesan Guruswami, Jason D. Hartline, Anna R. Karlin, David Kempe, Claire Kenyon, and Frank McSherry. On profit-maximizing envy-free pricing. In *Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1164–1173, 2005.

[GJ79]   Michael R. Garey and David S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W.H. Freeman, 1979.

[HN12]   Sergiu Hart and Noam Nisan. Approximate revenue maximization with multiple items. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, pages 656–656, 2012.

[MMW89]   R. Preston McAfee, John McMillan, and Michael D. Whinston. Multiproduct monopoly, commodity bundling, and correlation of values. *The Quarterly Journal of Economics*, 104(2):371–383, 1989.

[MV07]   Alejandro M. Manelli and Daniel R. Vincent. Multidimensional mechanism design: Revenue maximization and the multiple-good monopoly. *Journal of Economic Theory*, 137(1):153–185, 2007.

[Mye81]   Roger B. Myerson. Optimal auction design. *Mathematics of Operations Research*, 6(1):58–73, 1981.

[PP11]   Christos H. Papadimitriou and George Pierrakos. On optimal single-item auctions. In *Proceedings of the 43rd Annual ACM Symposium on Theory of Computing*, pages 119–128, 2011.

[Wil96]   Robert Butler Wilson. Nonlinear pricing and mechanism design. In H.M. Amman, D.A. Kendrick, and J. Rust, editors, *Handbook of Computational Economics*, volume 1, chapter 5, pages 253–293. Elsevier, 1996.