

Mechanism Design: From Optimal Transport Theory to Revenue Maximization

by

Christos Tzamos

Submitted to the Department of Electrical Engineering and Computer
Science

in partial fulfillment of the requirements for the degree of

PhD in Computer Science and Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2017

© Massachusetts Institute of Technology 2017. All rights reserved.

Author
Department of Electrical Engineering and Computer Science
May 17, 2017

Certified by
Constantinos Daskalakis
Associate Professor
Thesis Supervisor

Accepted by
Leslie A. Kolodziejcki
Chairman, Department Committee on Graduate Students

Mechanism Design: From Optimal Transport Theory to Revenue Maximization

by

Christos Tzamos

Submitted to the Department of Electrical Engineering and Computer Science
on May 17, 2017, in partial fulfillment of the
requirements for the degree of
PhD in Computer Science and Engineering

Abstract

A central problem in Economics and Algorithmic Game Theory is the design of auctions that maximize the auctioneer's expected revenue. While optimal selling of a single item has been well-understood since the pioneering work of Myerson in 1981, extending his work to multi-item settings has remained a challenge. In this work, we obtain such extensions providing a mathematical framework for finding optimal mechanisms.

In the first part of the work, we study revenue maximization in single-bidder multi-item settings, connecting this problem to a well-studied problem in measure theory, namely the design of optimal transport maps. By establishing strong duality between these two problems, we obtain a characterization of the structure of optimal mechanisms. As an important application, we prove that a grand bundling mechanism is optimal if and only if two measure-theoretic inequalities are satisfied. Likewise, we obtain necessary and sufficient conditions for the optimality of any mechanism in terms of a collection of measure-theoretic inequalities. Using our machinery we derive closed-form solutions in several example scenarios, illustrating the richness of mechanisms in multi-item settings, and we prove that the mechanism design problem in general is computationally intractable even for a single bidder.

In the second part of the work, we study multi-bidder settings where bidders have uncertainty about the items for sale. In such settings, the auctioneer may wish to reveal some information about the item for sale in addition to running an auction. While prior work has focused only on the information design part keeping the mechanism fixed, we study the combined problem of designing the information revelation policy together with the auction format. We find that prior approaches to this problem are suboptimal and identify the optimal mechanism by connecting this setting to the multi-item mechanism design problem studied in the first part of the work.

Thesis Supervisor: Constantinos Daskalakis
Title: Associate Professor

Acknowledgments

Foremost, I would like to thank my advisor, Costis Daskalakis, for his belief in me and his constant help and support through my years at MIT. Costis was a great advisor and collaborator, who taught me to seek important research questions with wide impact and work hard to answer them. I was very fortunate to have him as a mentor and working with him on a wide range of challenging problems from Economics to Learning Theory. He motivated me to look at the essence of things and his guidance was crucial in my development as a researcher. I am deeply grateful for the chance he gave me to collaborate with him and for his trust and support.

A big thank you goes also to Christos Papadimitriou. His contribution was essential for a big part of this dissertation and it was a great experience working with him. Moreover, I would also like to thank Dimitris Fotakis for introducing me to Theoretical Computer Science during my studies at NTUA and his guidance through the early steps of my career. I am thankful to Anindya De for our collaboration, Dirk Bergemann for his very helpful suggestions and discussions about several parts of this work as well as Silvio Micali and Parag Pathak for serving on my dissertation committee.

I would also like to thank the members of the theory group at MIT for the insightful discussions about research questions as well as for the fun we had together during my time at MIT. Special thanks to my collaborators Alan Deckelbaum, Matt Weinberg, Gautam Kamath, Manolis Zampetakis, Nishanth Dikkala, Themis Gouleakis, Arturs Backurs, Dimitris Tsipras and Kyriakos Axiotis for the many exciting hours we spent thinking together!

Last but not least, I am grateful for the support of my family. My wife, my parents and my brother are always close to me and patiently stand by me through the difficult times in my life.

Contents

| | | |
|----------|---|-----------|
| 1 | Introduction | 15 |
| 1.1 | Selling Multiple Items | 16 |
| 1.2 | Handling Bidder Uncertainty | 18 |
| 1.3 | Organization of the Dissertation | 19 |
| I | Designing Multi-Item Mechanisms | 21 |
| 2 | Overview and Related Work | 23 |
| 2.1 | Our results | 24 |
| 2.1.1 | Strong Duality | 25 |
| 2.1.2 | Characterization of Optimal Mechanisms | 27 |
| 2.1.3 | Complexity of Optimal Mechanism Design | 28 |
| 2.2 | Related Work | 30 |
| 3 | Problem Formulation | 35 |
| 3.1 | Revenue Maximization as Optimization | 35 |
| 3.1.1 | The Optimal Mechanism Design Problem | 35 |
| 3.1.2 | An Alternative Formulation | 36 |
| 3.1.3 | Example | 39 |
| 4 | A Dual to Mechanism Design | 41 |
| 4.1 | A Mechanism Design Strong Duality Theorem | 41 |
| 4.1.1 | Measure-Theoretic Preliminaries | 43 |

| | | |
|----------|--|-----------|
| 4.1.2 | Mechanism Design Duality | 44 |
| 4.2 | Single-Item Applications and Interpretation | 47 |
| 4.3 | Multi-Item Applications of Duality | 49 |
| 4.3.1 | Two Uniform $[0, 1]$ Items | 49 |
| 4.3.2 | Two Uniform But Not Identical Items | 51 |
| 4.3.3 | Discussion | 55 |
| 4.4 | Proof of Strong Mechanism Design Duality | 56 |
| 4.4.1 | A Strong Duality Lemma | 56 |
| 4.4.2 | From Two Convex Functions to One | 60 |
| 4.4.3 | Proof of Theorem 2 | 62 |
| 5 | Characterization of Optimal Mechanisms | 65 |
| 5.1 | Optimal Finite-Menu Mechanisms | 65 |
| 5.2 | Grand Bundling Optimality and Applications | 70 |
| 5.2.1 | Application I: Selling items with large values | 70 |
| 5.2.2 | Application II: Selling a large number of items | 76 |
| 5.3 | Proof of Stochastic Conditions of Section 5.1 | 78 |
| 5.3.1 | Probabilistic Lemmas | 78 |
| 5.3.2 | Proof of the Optimal Menu Theorem (Theorem 3) | 81 |
| 6 | A Framework for Computing Optimal Mechanisms | 89 |
| 6.1 | Constructing Optimal Mechanisms | 89 |
| 6.1.1 | Constructing Optimal Mechanisms for 2 Items | 91 |
| 6.2 | Verifying Stochastic Dominance | 95 |
| 6.2.1 | Proof of Lemma 11 | 97 |
| 6.3 | Application: An Optimal Mechanism with an Infinite Menu Size | 107 |
| 6.3.1 | Uniqueness of Mechanism in Example 3 | 113 |
| 6.4 | More Applications: Distributions of Unbounded Support | 115 |
| 6.4.1 | An Example with Power-Law Distributions | 116 |
| 6.4.2 | An Example with Exponential Distributions | 118 |

| | | |
|-----------|--|------------|
| 7 | Complexity of Mechanism Design | 121 |
| 7.1 | The Optimal Mechanism Design Problem | 122 |
| 7.1.1 | Our Approach | 128 |
| 7.2 | A Simple Class of Instances | 129 |
| 7.3 | A Linear Programming Approach | 131 |
| 7.3.1 | Mechanism Design as a Linear Program | 131 |
| 7.3.2 | A Relaxed Linear Program | 131 |
| 7.3.3 | The Min-Cost Flow Dual of the Relaxed Program | 134 |
| 7.4 | Characterizing Linear Programming Solutions | 134 |
| 7.4.1 | The Canonical Solution to LP3 | 135 |
| 7.4.2 | From LP3 to LP2 Solutions | 135 |
| 7.4.3 | From LP2 to LP1 Solutions | 136 |
| 7.4.4 | Putting it All Together | 138 |
| 7.5 | Proof of Theorem 8 | 139 |
| 7.6 | Beyond Additive or Computationally Bounded Bidders | 145 |
| 7.6.1 | Budget-Additive Bidders | 145 |
| 7.6.2 | Further Discussion: Powerful Bidders | 147 |
| II | Two-Sided Information Asymmetry | 151 |
| 8 | Overview and Related Work | 153 |
| 8.1 | Our Model and Results | 158 |
| 9 | Problem Formulation | 163 |
| 9.1 | A General Auction Model | 163 |
| 9.2 | An Important Special Case | 164 |
| 9.3 | Signaling Schemes | 165 |
| 9.4 | Auction Formats | 166 |
| 10 | Handling Bidder Uncertainty | 167 |
| 10.1 | Optimal Auctions | 167 |

| | |
|---|------------|
| 10.1.1 Myerson (No Information Revelation) | 167 |
| 10.1.2 Myerson with Signaling | 167 |
| 10.1.3 Optimal Augmented Auction | 170 |
| 10.2 Connections to Multi-Item Mechanisms | 174 |
| 11 Revenue Comparison | 177 |
| 11.1 Single Bidder | 178 |
| 11.2 Multiple Bidders | 179 |
| 12 Discussion and Extensions | 183 |
| 12.1 Multiple-Items | 183 |
| 12.2 Structure and Computation | 184 |
| 12.3 Correlation Among Bidders and Item Types | 185 |
| 13 Conclusions and Future Directions | 189 |

List of Figures

| | | |
|-----|---|-----|
| 4-1 | The MV-mechanism for two i.i.d. uniform $[0, 1]$ items. | 50 |
| 4-2 | Partition of $[4, 16] \times [4, 7]$ into different regions by the optimal mechanism. | 52 |
| 5-1 | Partition of the type set $X = [0, 100]^2$ induced by some menu of lotteries. | 66 |
| 5-2 | The regions of Lemma 4 for the case $n = 3$ | 71 |
| 6-1 | The canonical partition | 92 |
| 6-2 | We show that either decreasing $z_2^{(k+1)}$ to $z_2^{(k)}$ or removing $z^{(k+1)}$ entirely decreases the value of $\int_A (f - g)$. In either case, we can apply our inductive hypothesis. | 106 |
| 6-3 | The well-formed canonical partition for $f_1(x) = 2(1 - x)$ and $f_2(y) = 2(1 - y)$ | 110 |
| 6-4 | Canonical Partitions for different cases of Beta distributions. The shaded region is where the measure μ becomes negative. (Note that when the second parameter b_i of the Beta distribution of some item i equals 1, μ has positive mass on the outer boundary $x_i = 1$.) (1) Beta(1,1) and Beta(1,1), (2) Beta(2,2) and Beta(1,1), (3) Beta(2,2) and Beta(2,2). | 112 |
| 6-5 | Decomposition of measure $\mu _{\mathcal{W}}$ into measures $\mu' _{\mathcal{W} \setminus R'}$ and $\mu _{\mathcal{W}} - \mu' _{\mathcal{W} \setminus R'}$. The dark shaded regions R' and $\mathcal{H} = [1 - \epsilon, 1]^2$ show the support of $\mu _{\mathcal{W}} - \mu' _{\mathcal{W} \setminus R'}$ | 115 |

| | | |
|-----|--|-----|
| 6-6 | The canonical partition of $\mathbb{R}_{\geq 0}^n$ for the proof of Example 5. In this diagram, $p^* > 2/\lambda_1$. If $p^* \leq 2/\lambda_1$, \mathcal{B} is empty. The positive part μ_+ of μ is supported inside $\mathcal{P} \cap \{\vec{0}\}$ while the negative part μ_- is supported within $Z_{p^*} \cup \mathcal{N}$ | 119 |
| 7-1 | LP1, the linear program for revenue maximization | 131 |
| 7-2 | LP2, the relaxed linear program | 133 |
| 7-3 | LP3, the dual of LP2 | 134 |

List of Tables

| | |
|--|-----|
| 11.1 Comparison of the worst-case approximation gap among the different auction formats for the case of a single bidder | 178 |
| 11.2 Comparison of the worst-case approximation gap among the different auction formats for the case of many bidders | 180 |

Chapter 1

Introduction

The problem of optimal mechanism design has been a central problem in Economics for decades. With the emergence and prevalence of online ad auctions it has become a defining problem of the modern economy and has also been extensively studied by Computer Science. In the most basic version of this problem, a seller has a single item to sell to a group of interested buyers. Each buyer knows his own value for the item, but the seller and potentially the other buyers only know a distribution from which these values are assumed to be drawn. The goal is to design a sales procedure, called a *mechanism*, that optimizes the expected revenue of the seller.

The problem of optimal mechanism design and the theory behind it find a large number of applications in various areas. When auction-houses sell items, this is the problem that they face. This is also the problem that governments face when auctioning a valuable public resource such as wireless spectrum. Finally, the problem arises every millisecond as auctions are used in sponsored search and the allocation of banner advertisements.

In the basic form of the problem of selling a single item, optimal mechanism design is really well understood. Building on Myerson's celebrated work [Mye81], it has been studied intensely for decades in both Economics and Computer Science. This research has revealed surprisingly elegant structure in the optimal mechanism, as well as robustness to the details of the distributions, and has had a deep impact in the broader field of mechanism design. In particular, in the case of one buyer selling

the item at a predetermined price is optimal while in the case of many buyers, the optimal mechanism is a second price auction with a reserve price. More precisely,

Fact 1 ([Mye81]). *An optimal way to sell one item to a buyer whose value for the item is drawn from some known distribution F is to make a take-it-or-leave-it offer at a price p^* in $\arg \max\{x \cdot (1 - F(x))\}$.*

For many buyers whose values are drawn independently from the same regular¹ distribution F , an optimal mechanism is a second-price auction with reserve price p^ .*

The structure of an optimal mechanism is also known for arbitrary distributions and when bidders are non i.i.d.

While all this progress has been taking place on the single-item front, our understanding beyond this simple setting is still significantly limited. In this work, we focus two important and natural extensions that find applications in many real world scenarios: selling multiple-items and handling buyer uncertainty.

1.1 Selling Multiple Items

Many of the aforementioned applications, such as allocating sponsored search advertisements or distributing frequencies in the wireless spectrum, are inherently multi-item. Unfortunately, our understanding of how to optimize revenue in such settings is severely limited.

Despite substantial research effort, it is not even known how to optimally sell two items to one buyer with additive valuations. It is worth mentioning that the optimal multi-item mechanism might not be comprised of optimal single-item mechanisms, even when the values of the items are distributed independently. On the contrary, multi-item auctions appear to have a much richer structure, often exhibiting unintuitive properties. Even when the item values are independent, the mechanism may benefit from selling bundles of items or even lotteries over bundles of items [MMW89, BB99, Tha04, MV06]. Moreover, no general framework to approach

¹A distribution F with density f is called *regular* iff $x - \frac{1-F(x)}{f(x)}$ is monotone increasing.

this problem has been proposed in the literature, making it dauntingly difficult both to identify optimal solutions and to certify the optimality of those solutions. As a consequence, seemingly simple special cases (even with two items) remain poorly understood, despite much research for a few decades.

One of the key contributions of our work is a structural characterization of optimal multi-item auctions. We find that the revenue maximization problem has a tight dual problem which takes the form of a well-studied problem in Mathematics, the problem of optimal mass transportation. We exploit this connection to understand when mechanisms that sell all item together for a single price are optimal. We show that such mechanisms, called *grand-bundling mechanisms*, are optimal if and only if two stochastic dominance conditions hold between specific measures induced by the buyer’s value distribution. This answers a longstanding question of under which conditions grand bundling is optimal. The necessary and sufficient condition for grand bundling optimality is a special case of our more general characterization result that provides necessary and sufficient conditions for the optimality of an arbitrary mechanism (with a finite menu size) for an arbitrary type distribution.

Building on our characterization, we provide a framework novel framework for revenue maximization and identify optimal mechanisms in a wide range of different settings. Our framework is applicable to arbitrary number of items and distributions, with mild assumptions such as differentiability. In particular, we strengthen prior work [MV06, DDT13, GK14], which identified optimal mechanisms in special cases. We exhibit the practicality of our framework by solving several examples and observe that even for instances with two items and one buyer whose values for the items are independently drawn from very simple distributions such as beta distributions, the resulting revenue-optimal mechanism offers an uncountably large number of different bundles to the buyer to choose from.

Complementing our framework for identifying optimal mechanisms for continuous distributions, we show that designing revenue-optimal auctions is computationally intractable for large number of items, even in quite simple settings involving a single additive buyer whose values for every item are independently drawn from distributions

of support 2.

1.2 Handling Bidder Uncertainty

Another important extension to the basic formulation of optimal mechanism design is modelling the potential uncertainty that buyers have for the item for sale. Even in a single buyer, single item setting, the buyer's value might depend on features of the item that are unobservable to him at the point of sale. We call this a *two sided information asymmetry* setting, as opposed to the single sided information asymmetry setting of the basic model, since neither the seller nor the buyer knows exactly how much the item is worth. Indeed, the buyer doesn't observe the item's realized type, which affects his value, while the seller does not observe the buyer's realized type, which affects his value for the realized type of the item.

For example, when buying clothes online a buyer's value might depend on attributes such as material feel, shape and fit which will only be known to the buyer after he buys the item. As another example, in the ad-auctions applications we saw earlier, an advertiser's value for a click depends on several properties of the user clicking on the ad, such as age, gender or location, which are often unobservable to the advertiser. In such cases, the ad platform which might have a lot more information about the user might want to reveal information to the advertiser to increase its expected revenue.

The main question we would like to study in this work is how much information the seller should reveal to the buyers to maximize his expected revenue. Such a question can be asked with a fixed mechanism in mind where the goal is to find the optimal amount of information to reveal prior to running the mechanism. More importantly, though, we are interested in a combined design problem where we have the flexibility to design a mechanism that specifies both a information revelation policy and an allocation policy (potentially multi-round) to be used by the auctioneer. We call such mechanisms "augmented auctions".

Our main result is a one-to-one correspondence between any setting with two-sided

information asymmetry and an associated multi-item setting with one-sided information asymmetry, i.e. an instance of the multi-item setting discussed earlier. We find that multi-item mechanisms directly correspond to equivalent augmented auctions. In particular, selling multiple items separately corresponds to fully revealing all available information and running Myerson’s optimal auction [Mye81] which is described in Fact 1. Additionally, selling multiple items in a grand-bundling mechanism corresponds to revealing no information whatsoever and running Myerson’s optimal auction. And more importantly, the optimal multi-item auction directly corresponds to and yields the optimal augmented auction.

1.3 Organization of the Dissertation

The dissertation consists of two parts. In the first part, we study multi-item mechanisms, while in the second part we study mechanisms with two-sided information asymmetry. The first part is based on joint work with Constantinos Daskalakis and Alan Deckelbaum [DDT13, DDT14, DDT17], while the second part is based on joint work with Constantinos Daskalakis and Christos Papadimitriou [DPT16].

Each part begins with an overview chapter (Chapters 2 and 8) that presents our main results in more detail and provides details and pointers to related work. Then, we provide an additional chapter in each part where we formally state the problem and provide the necessary notation and background needed in the proof of our results.

For the first part, we focus on designing optimal multi-item mechanisms:

- In Chapter 4, we provide and prove our main analytical tool that obtains a strong dual to the multi-item mechanism design problem. The dual takes the form an optimal transportation problem. More generally, we provide necessary and sufficient conditions for arbitrary mechanisms with a finite menu.
- In Chapter 5, we build on our strong duality to obtain a characterization of optimal auctions. We apply our characterization to give necessary and sufficient conditions for optimality of grand bundling mechanisms.

- In Chapter 6, we use our strong duality theorem and characterization of optimal auctions to provide a framework for identifying optimal mechanisms and give several examples for different closed form distributions.
- In Chapter 7, we show that the problem of computing the revenue optimal multi-item mechanism is computationally intractable even for the case of a single additive bidder with independent valuations.

For the second part, we focus on designing optimal mechanisms for settings with bidder uncertainty:

- In Chapter 10, we connect the setting with bidder uncertainty to the setting with multiple items where bidders have full information about the items for sale. We show how mechanisms for one directly correspond to mechanisms for the other setting through an extension of the revelation principle.
- In Chapter 11, we show that mechanisms that exploit the information asymmetry can achieve significantly higher revenue than mechanisms that fully reveal all relevant information to the buyers.
- In Chapter 12, we discuss extensions and implications of the connection to multi-item auctions. We exploit this connection to multi-item auctions to readily obtain efficient algorithms for computing optimal auctions with two-sided information asymmetry.

Finally, in Chapter 13, we conclude with a summary of our work and a discussion of open problems and future research.

Part I

Designing Multi-Item Mechanisms

Chapter 2

Overview and Related Work

We study the problem of revenue maximization for a multiple-good monopolist. Given n heterogenous goods and a probability distribution f over $\mathbb{R}_{\geq 0}^n$, we wish to design a mechanism that optimizes the monopolist’s expected revenue against an additive (linear) buyer whose values for the goods are distributed according to f .

The single-good version of this problem—namely, $n = 1$ —is well-understood, going back to [RS81, Mye81, MR84, RZ83], where it is shown that a take-it-or-leave-it offer of the good at some price is optimal, and the optimal price can be easily calculated from f (see Fact 1). For general n , it has been known that the optimal mechanism may exhibit much richer structure.

To gain some intuition, it is worth pointing out that the optimal mechanism for selling multiple items to an additive bidder is not necessarily comprised of the optimal mechanisms for selling each item separately. Even with two items whose values are independent, the mechanism may benefit from selling bundles of items or even lotteries over bundles of items [MMW89, BB99, Tha04, MV06].

Here is an example from Hart and Nisan [HN12]. Suppose that there are two items and the bidder’s value for each is either 1 or 2 with probability $\frac{1}{2}$, independently of the other item. In this case, the maximum expected revenue from selling the items separately is 2, achieved e.g. by posting a price of 1 on each item. However, offering instead the bundle of both items at a price of 3 achieves a revenue of $3 \cdot \frac{3}{4} = \frac{9}{4}$.

On the other hand, bundling the items together is not always better than selling

them separately. If there are two items with values 0 or 1 with probability $\frac{1}{2}$, independently from each other, then selling the bundle of the two items achieves revenue at most $\frac{3}{4}$, but selling the items separately yields revenue of 1, which is optimal.

In general, the optimal mechanism may have much more intricate structure than either selling the grand bundle of all the items, or selling each item separately, even when the item values are i.i.d. In fact, the optimal mechanism might not even be deterministic: we may need to offer for sale not only sets of items, but also lotteries over sets of items. Here is an example. Suppose that there are two items, one of which is distributed uniformly in $\{1, 2\}$ and the other uniformly in $\{1, 3\}$, and suppose that the items are independent. In this case, the optimal mechanism offers the bundle of the two items at price 4, and it also offers at price 2.5 a lottery that with probability $\frac{1}{2}$ gives both the items and with probability $\frac{1}{2}$ just the first item. The optimality of the mechanism follows from our techniques of Section 7.4.4. Finally, consider an example with 2 items where the item values are drawn independently from Beta(1,2). As we show in Section 6.3, the optimal mechanism in such settings may offer an uncountably large menu of randomized bundles to the buyer.

As these examples illustrate, optimal mechanisms can have a very rich behavior. Moreover, no general framework to approach this problem has been proposed in the literature, making it dauntingly difficult both to identify optimal solutions and to certify the optimality of those solutions. As a consequence, seemingly simple special cases (even $n = 2$) remain poorly understood, despite much research for a few decades.

2.1 Our results

We propose a novel framework for revenue maximization based on duality theory. We identify a minimization problem that is dual to revenue maximization and prove that the optimal values of these problems are always equal. The dual corresponds to a well-studied problem in Mathematics, the problem of designing optimal transportation maps. Our framework allows us to identify optimal mechanisms in general settings, and certify their optimality by providing a complementary solution to the

dual problem, namely finding a solution to the dual whose objective value equals the mechanism’s revenue. Our framework is applicable to arbitrary settings of n and f , with mild assumptions such as differentiability. In particular, we strengthen prior work [MV06, DDT13, GK14], which identified optimal mechanisms in special cases. We exhibit the practicality of our framework by solving several examples. Importantly, we can leverage our duality theorem to characterize optimal multi-item mechanisms. From a technical standpoint we provide new analytical methodology for multi-dimensional mechanism design by providing extensions to Monge-Kantorovich duality for optimal transportation. We proceed to discuss our contributions in detail, providing a roadmap to the results, and conclude this section with a discussion of related work.

2.1.1 Strong Duality

Our first main result (presented as Theorem 2) formulates a dual problem to the optimal mechanism design problem, and establishes strong duality between the two problems. That is, we show that the optimal values of the two optimization problems are identical. Our approach for developing this dual problem is outlined below.

We start by formulating optimal mechanism design as a maximization problem over convex, non-decreasing and 1-Lipschitz continuous functions u , representing the utility of the buyer as a function of his type, as in [Roc87]. The objective function of this maximization problem can be written as the expectation of u with respect to a signed measure μ over the type space of the buyer. Measure μ is easily derived from the buyer’s type distribution f (see Equation (3.3)) and expresses the marginal change in the seller’s revenue under marginal changes in the rent paid to subsets of buyer types. Our formulation is summarized in Theorem 1, while Section 3.1.3 illustrates our formulation in the basic setting of independent uniform items.

In Theorem 2, we formulate a dual in the form of an optimal transportation problem, and establish strong duality between the two problems. Roughly speaking, our dual formulation is given the signed measure μ (from Theorem 1) and solves the following minimization problem: (i) first, it is allowed to choose any measure μ' that

stochastically dominates μ with respect to convex increasing functions; (ii) second, it is supposed to find a coupling of the positive part μ'_+ of μ' with its negative part μ'_- i.e. find a transportation from μ'_+ to μ'_- ; (iii) if a unit of mass of μ'_+ at x is transported to a unit of mass of μ'_- at y , we are charged $\|x - y\|_1$. The goal is to minimize the cost of the coupling with respect to the decisions in (i) and (ii).

While our dual formulation takes a simple form, establishing strong duality is quite technical. At a high level, our proof follows the proof of Monge-Kantorovich duality in [Vil08], making use of the Fenchel-Rockafellar duality theorem, but the technical aspects of the proof are different due to the convexity constraint on feasible utility functions. We note that our formulation from Theorem 1 defines a convex optimization problem. One would hope then that infinite-dimensional linear programming techniques [Lue68, AN87] can be leveraged to establish the existence of a strong dual. We are not aware of such an approach, and expect that such formulations will fail to establish existence of interior points in the primal feasible set, which is necessary for strong duality.

As already emphasized earlier, our identification of a strong dual implies that the optimal mechanism admits a certificate of optimality, in the form of a dual witness, for all settings of n and f . Hence, our duality framework can play the role of first-order conditions certifying the optimality of single-dimensional mechanisms. Where optimality of single-dimensional mechanisms can be certified by checking virtual welfare maximization, optimality of multi-dimensional mechanisms is always certifiable by providing dual solutions whose value matches the revenue of the mechanism, and such dual solutions take a simple form: they are transportation maps between measures.

Using our framework, we can provide shorter proofs of optimality of known mechanisms. As an illustrating example, we show in Section 4.3.1 how to use our framework to establish the optimality of the mechanism for two i.i.d. uniform $[0, 1]$ items proposed by [MV06]. Then in Section 4.3.2, we provide a simple illustration of the power of our framework, obtaining the optimal mechanism for two independent uniform $[4, 16]$ and uniform $[4, 7]$ items, a setting where the results of [MV06, Pav11, DDT13,

[GK14] fail to apply. The optimal mechanism has the somewhat unusual structure shown in the diagram in Section 4.3, where types in Z are allocated nothing (and pay nothing), types in W are allocated the grand bundle (at price 12), while types in Y are allocated item 2 and get item 1 with probability 50% (at price 8).

2.1.2 Characterization of Optimal Mechanisms

Substantial effort in the literature has been devoted to studying optimality of mechanisms with a simple structure such as pricing mechanisms; see, e.g., [MV06] and [DDT13] for sufficient conditions under which mechanisms that only price the grand bundle of all items are optimal. Our second main result (presented as Theorem 3) obtains *necessary and sufficient* conditions characterizing the optimality of arbitrary mechanisms with a finite menu size. We proceed to describe our characterization result in more detail.

Suppose that we are given a feasible mechanism \mathcal{M} whose set of possible allocations is finite. We can then partition the type set into finitely many subsets (called regions) $\mathcal{R}_1, \dots, \mathcal{R}_k$ of types who enjoy the same price and allocation. The question is this: *for what type distributions is \mathcal{M} optimal?* Theorem 3 answers this question with a sharp characterization result: \mathcal{M} is optimal *if and only if* the measure μ (derived from the type distribution as described above) satisfies k stochastic dominance conditions, one per region in the afore-defined partition. The type of stochastic dominance that μ restricted to region \mathcal{R}_i ought to satisfy depends on the allocation to types from \mathcal{R}_i , namely which set of items are allocated with probability 1, 0, or non-0/1.

Theorem 3 is important in that it reduces checking the optimality of mechanisms to checking standard stochastic dominance conditions between measures derived from the type distribution f , which is a concrete and easier task than arguing optimality against all possible mechanisms.

Theorem 3 is a corollary of our strong duality framework (Theorem 2), but requires a sequence of technical results. One direction of our characterization result requires turning the stochastic dominance conditions into dual solutions that can be plugged into Theorem 2 to establish the optimality of a given mechanism. The other direction

requires showing that a dual solution certifying the optimality of a given mechanism also implies that the stochastic dominance conditions of Theorem 3 must hold.

A particularly simple special case of our characterization result pertains to the optimality of the grand-bundling mechanism. See Theorem 4. We show that the mechanism offering the grand bundle at price p is optimal *if and only if* measure μ satisfies a pair of stochastic dominance conditions. In particular, if Z are the types who cannot afford the grand bundle and W the types who can, then offering the grand bundle for p is optimal if and only if the following conditions hold:

- $\mu_{-|Z}$, the negative part of μ restricted to Z , stochastically dominates $\mu_{+|Z}$, the positive part of μ restricted to Z , with respect to all convex increasing functions;
- $\mu_{+|W}$ stochastically dominates $\mu_{-|W}$ with respect to all concave increasing functions.

Already our characterization of grand-bundling optimality settles a long line of research which only obtained sufficient conditions for the optimality of grand-bundling.

In turn, we illustrate the power of our characterization of grand-bundling optimality with Theorems 5 and 6, two results that are interesting on their own right. Theorem 5 generalizes the corresponding result of [Pav11] from two to an arbitrary number of items. We show that, for any number of items n , there exists a large enough c such that the optimal mechanism for n i.i.d. uniform $[c, c + 1]$ items is a grand-bundling mechanism. While maybe an intuitive claim, we do not see a direct way of proving it. Instead, we utilize Theorem 4 and construct intricate couplings establishing the stochastic dominance conditions required by the theorem. In view of Theorem 5, our companion theorem, Theorem 6, seems even more surprising. We show that in the same setting of n i.i.d. uniform $[c, c + 1]$ items, for any fixed c it holds that, for all sufficiently large n , the optimal mechanism *is not* (!) a grand-bundling mechanism. See Section 5.1 for the proofs of these results.

2.1.3 Complexity of Optimal Mechanism Design

Motivated by the richness in behavior of the optimal multi-item mechanisms, we study the complexity of computing them. We show that this multi-item design problem is

computationally intractable, even in the most basic settings.

We show that there is no expected polynomial-time solution to the optimal mechanism design problem unless $ZPP \supseteq P^{\#P}$. In particular, it is $\#P$ -hard to determine whether every optimal mechanism assigns a specific item to a specific type of bidder with probability 0 or with probability 1, given the promise that one of these two cases holds simultaneously for all optimal mechanisms. Our complexity result holds even in the case of a single additive, quasi-linear bidder, whose values for the items are independently distributed on two rational numbers with rational probabilities.

We present a formal definition of the optimal mechanism design problem in Section 7.1. In the single bidder setting, any efficient solution enables us, given the bidder’s declared type, to exactly sample from the appropriate allocation rule and to sample a randomized payment with the appropriate expected value.¹ In our hardness proof, a single sample from the allocation distribution would enable us to distinguish whether the bidder should receive a certain item with probability 0 or with probability 1. We note that our proof is subject to the assumption $ZPP \not\supseteq P^{\#P}$ (rather than $P \not\supseteq P^{\#P}$) solely because we prove lower bounds for randomized mechanisms.

The contribution of our complexity result is two-fold. First, it gives a definitive proof that approximation is *necessary* for revenue optimization beyond Myerson’s single-item setting, due to *computational considerations*, even in the ideal scenario where the value distributions are perfectly known. Approximation has been heavily used in algorithmic work on the problem, but there has been no justification for its use, at least in simple settings that don’t induce combinatorial structure in the valuations of the bidders (sub-modular, OXS, etc.), or the allocation constraints of the setting. Second, our result represents advancement in our techniques for showing lower bounds for optimal mechanism design. Despite evidence that the structure of the optimal mechanism is complicated even in simple settings, previous work has not been able to harness this evidence to obtain computational hardness results. Again we note that complexity creeps in not because we assume correlations in the item values (which

¹Since any computable payment function can be realized as the expectation of an expected polynomial-time computable random variable (with possibly huge maximum value), to prove our hardness result we focus on the difficulty of sampling from the allocation rule.

can easily introduce combinatorial dependencies among them), or because we restrict attention to deterministic mechanisms (whose structure is readily combinatorial), but because our approach reveals combinatorial structure in the optimal mechanism, which can be exploited for a reduction.

2.2 Related Work

There is a rich literature on multi-item mechanism design pertaining to the multiple good monopoly problem that we consider here. We refer the reader to the surveys [RS03, MV07, FKM11] for a detailed description, focusing on the work closest to ours.

Much work has focused on obtaining sufficient conditions for optimality of mechanisms. Hart and Nisan [HN12], Menicucci et al [MHJ15] and Haghpanah and Hartline [HH15] provide sufficient conditions for the grand-bundling mechanism to be optimal. Manelli and Vincent [MV06] provide conditions for the optimality of more complex deterministic mechanisms and, similarly, [DDT13, GK14] provide sufficient conditions for the optimality of general (possibly randomized) mechanisms. Finally, Haghpanah and Hartline [HH15] provide an approach for reverse engineering sufficient conditions for a simple mechanism to be optimal. These works on sufficient conditions apply to limited settings of n and f . They typically proceed by relaxing some of the truthfulness constraints and are therefore only applicable when the relaxed constraints are not binding at the optimum.

In addition to sufficient conditions, a lot of work has focused on characterizing structural properties of optimal mechanisms. Armstrong [Arm96] has shown that optimal mechanisms always exclude a fraction of buyer types of low value from the mechanism. Thanassoulis [Tha04], Briest et al [BCKW10] and Hart and Nisan [HN13] show that randomization is necessary for optimal revenue extraction. In turn, Manelli and Vincent [MV07] have shown that there exist type distributions for which optimal mechanisms are arbitrarily complex. Moreover, Hart and Reny [HR15] provide an interesting example where a product type distribution over two items stochastically

dominates another, yet the optimal revenue from the weaker distribution is higher.

Rochet and Choné [RC98] study a closely related setting, providing a characterization of the optimal mechanism for the multiple good monopoly problem where the monopolist has a (strictly) convex cost for producing copies of the goods. With strictly convex production costs, optimal mechanism design becomes a strictly concave maximization problem, which allows the use of first-order conditions to characterize optimal mechanisms. Our problem can be viewed as having a production cost that is 0 for selling at most one unit of each good and infinity otherwise. While still convex, our production function is not strictly convex and is discontinuous, making first-order conditions less useful for characterizing optimal mechanisms. This motivates the use of duality theory in our setting. From a technical standpoint, optimal mechanism design necessitates the development of new tools in optimal transport theory [Vil08], extending Monge-Kantorovich duality to accommodate convexity constraints in the dual of the transportation problem. In our setting, the dual of the transportation problem corresponds to the mechanism design problem and these constraints correspond to the requirement that the utility function of the buyer be convex, which is intimately related to the truthfulness of the mechanism [Roc87]. In turn, accommodating the convexity constraints in the mechanism design problem requires the introduction of mean-preserving spreads of measures in its transportation dual, resembling the “multi-dimensional sweeping” of Rochet and Choné.

Ultimately, our work relies on and develops further a fundamental connection of optimal transportation to designing optimal mechanisms. See Ekeland’s notes on Optimal Transportation [Eke10] for more connections to mechanism design.

In addition to work on identifying and characterizing optimal mechanisms, some literature [Arm99, HN12, BILW14, LY13, CH13] has focused on the revenue guarantees of simple mechanisms for independent distributions, e.g. bundling all items together or selling them separately. Hart and Nisan [HN12] show that selling separately guarantees at least a $c/\log^2 n$ fraction of the revenue, which was later improved to a tight factor of $O(\log n)$ by Li and Yao [LY13]. Furthermore, Babaioff et al. [BILW14] showed that choosing to either bundle all items together or sell them separately guar-

antees a $1/6$ fraction of the revenue. Later, Yao [Yao15] extended this result to the case of many buyers providing a mechanism that achieves at least a constant fraction of the optimal revenue.

Significant progress has also been made in algorithmic solutions to optimal mechanism design problems. Cai et al. provide efficiently computable revenue-optimal [CDW12a, CDW12b] or approximately optimal mechanisms [CDW13] in very general settings, including settings where there are combinatorial constraints over which allocations of items to bidders are feasible. However, these results, as well as the more specialized ones of Alaei et al. [AFH⁺12] for service-constrained environments, apply to the explicit setting, i.e. when the distributions over bidders' valuations are given explicitly, by listing every valuation in their support together with the probability it appears. Such a manner of describing a joint distribution, however, is often unnatural. For example, in the case of independently distributed items, explicitly listing the joint distribution can be exponentially longer than listing each marginal distribution separately. In such a case, our complexity result implies that finding the optimal mechanism is computationally intractable even for simple independent distributions.

Prior to our work, there has been considerable effort towards computational lower bounds for optimal mechanism design. Nevertheless, all these results are for either somewhat exotic families of valuation functions or distributions over valuations, or for computing the optimal deterministic, as opposed to general (possibly randomized) auction. In the first vein, Dobzinski et al. [DFK11] show that optimal mechanism design for OXS bidders is NP-hard via a reduction from the CLIQUE problem. OXS valuations are described implicitly via a graph, include additive, unit-demand² and much broader classes of valuations, and are more amenable to lower bounds given the combinatorial nature of their definition. (See further discussion in Section 7.6.) In joint work with Daskalakis and Deckelbaum [DDT12], we prove #P-hardness of computing the optimal mechanism design problem for a single item and a single bidder whose value for the item is the sum of independently distributed attributes.

²A *unit-demand* bidder is described by a vector (v_1, \dots, v_n) of values, where n is the number of items. If the bidder receives item i , his value is v_i , while if he receives a set of more than one item, his value for that set is his value for his favorite item in the set.

Compared to these lower bounds, our goal is to prove intractability results for very simple valuations (namely additive) and simple distributions (namely the item values are independent and the distribution of each item is given explicitly), and which have no combinatorial structure incorporated in their definition. We prove our lower-bound for a single additive buyer, while follow-up work by Chen et al. [CDO⁺15] extends our results and techniques to the case of unit-demand buyers.

On the complexity of optimal deterministic auctions, when the bidder's values for the items are correlated according to some explicitly given distribution, Briest [Bri08] shows inapproximability results for selling multiple items to a single unit-demand bidder via an item-pricing auction that posts a price for each item, and lets the bidder buy whatever item he wants. More recently, Chen et al. [CDP⁺14] show hardness of computing optimal pricing mechanisms for unit-demand bidders under independent distributions. Papadimitriou and Pierrakos [PP11] show APX-hardness results for the optimal, incentive compatible, deterministic auction when a single item is sold to multiple bidders, whose values for the item are correlated according to some explicitly given distribution. (We note that the settings considered in Briest [Bri08], Papadimitriou-Pierrakos [PP11] and Chen et al. [CDP⁺14] are polynomial-time solvable via linear programming when the determinism requirement is removed.) Finally, together with Daskalakis and Deckelbaum [DDT12], we provide SQRTSUM-hardness results for the optimal item-pricing problem when there is a single unit-demand bidder whose values for the items are independent of support two, and when either the values or the probabilities may be irrational. Compared to these results, our lower bounds apply to general (i.e. possibly randomized) auctions, the values of the items are distributed independently, and both supports and probabilities are rational numbers.

Chapter 3

Problem Formulation

3.1 Revenue Maximization as Optimization

3.1.1 The Optimal Mechanism Design Problem

Our goal is to find the revenue-optimal mechanism \mathcal{M} for selling n goods to a single additive buyer. An additive buyer has a *type* x specifying his value for every good. The value of the buyer for a set of goods S is then simply equal to $\sum_{i \in S} x_i$. The type x of the buyer is an element of a *type space* $X = \prod_{i=1}^n [x_i^{\text{low}}, x_i^{\text{high}}]$, where $x_i^{\text{low}}, x_i^{\text{high}}$ are non-negative real numbers. While the buyer knows his type with certainty, the mechanism designer only knows the probability distribution over X from which x is drawn. We assume that the distribution has a density $f : X \rightarrow \mathbb{R}$ that is continuous and differentiable with bounded derivatives.

Without loss of generality, by the revelation principle [Mye79], we consider direct mechanisms. A (direct) mechanism consists of two functions: (i) an *allocation function* $\mathcal{P} : X \rightarrow [0, 1]^n$ specifying the probabilities, for each possible type declaration of the buyer, that the buyer will be allocated each good, and (ii) a *price function* $\mathcal{T} : X \rightarrow \mathbb{R}$ specifying, for each declared type of the buyer, the price that he is charged. When an additive buyer of type x declares himself to be of type $x' \in X$, he receives net expected utility $x \cdot \mathcal{P}(x') - \mathcal{T}(x')$.

We restrict our attention to mechanisms that are *incentive compatible*, meaning

that the buyer must have adequate incentives to reveal his values for the items truthfully, and *individually rational*, meaning that the buyer has an incentive to participate in the mechanism.

Definition 1. Mechanism $\mathcal{M} = (\mathcal{P}, \mathcal{T})$ over type space X is incentive compatible (IC) if and only if $x \cdot \mathcal{P}(x) - \mathcal{T}(x) \geq x \cdot \mathcal{P}(x') - \mathcal{T}(x')$ for all $x, x' \in X$.

Definition 2. Mechanism $\mathcal{M} = (\mathcal{P}, \mathcal{T})$ over type space X is individually rational (IR) if and only if $x \cdot \mathcal{P}(x) - \mathcal{T}(x) \geq 0$ for all $x \in X$.

The mechanism designer's goal is to find an incentive compatible and individually rational mechanism $\mathcal{M} = (\mathcal{P}, \mathcal{T})$ maximizing the expected revenue received by the mechanism, i.e. $\int_X \mathcal{T}(x) f(x) dx$. We call this the optimal mechanism design problem.

3.1.2 An Alternative Formulation

Instead of optimizing over mechanisms $\mathcal{M} = (\mathcal{P}, \mathcal{T})$ directly, we can rewrite the problem more succinctly in terms of only one function that expresses the utility the buyer enjoys in the mechanism. When a buyer truthfully reports his type to a mechanism $\mathcal{M} = (\mathcal{P}, \mathcal{T})$ (over type space X), we denote by $u : X \rightarrow \mathbb{R}$ the function that maps the buyer's valuation to the utility he receives by \mathcal{M} . It follows by the definitions of \mathcal{P} and \mathcal{T} that $u(x) = x \cdot \mathcal{P}(x) - \mathcal{T}(x)$.

It is well-known (see [Roc87], [RC98], and [MV06]), that an IC and IR mechanism has a convex, nonnegative, nondecreasing, and 1-Lipschitz utility function with respect to the ℓ_1 norm and that any utility function satisfying these properties is the utility function of an IC and IR mechanism with $\mathcal{P}(x) = \nabla u(x)$ and $\mathcal{T}(x) = \mathcal{P}(x) \cdot x - u(x)$.¹

We clarify that a function u is 1-Lipschitz with respect to the ℓ_1 norm if it satisfies $u(x) - u(y) \leq \|x - y\|_1$ for all $x, y \in X$. This is essentially equivalent to all partial derivatives having magnitude at most 1 in each dimension.

¹On the measure-0 set on which ∇u is not defined, we can use an analogous expression for \mathcal{P} by choosing appropriate values of ∇u from the subgradient of u .

We will now formulate the mechanism design problem for a mechanism $\mathcal{M} = (\mathcal{P}, \mathcal{T})$, directly as an optimization problem over feasible utility functions u . We first define the notation:

- $\mathcal{U}(X)$ is the set of all continuous, non-decreasing, and convex functions $u : X \rightarrow \mathbb{R}$.
- $\mathcal{L}_1(X)$ is the set of all 1-Lipschitz with respect to the ℓ_1 norm functions $u : X \rightarrow \mathbb{R}$.

In this notation, a mechanism \mathcal{M} is IC and IR if and only if its utility function u satisfies $u \geq 0$ and $u \in \mathcal{U}(X) \cap \mathcal{L}_1(X)$. It follows that the optimal mechanism design problem can be viewed as an optimization problem:

$$\sup_{\substack{u \in \mathcal{U}(X) \cap \mathcal{L}_1(X) \\ u \geq 0}} \int_X [\nabla u(x) \cdot x - u(x)] f(x) dx.$$

Notice that for any utility u defining an IC and IR mechanism, the function $\tilde{u}(x) = u(x) - u(x^{\text{low}})$ also defines a valid IC and IR mechanism since $\tilde{u} \in \mathcal{U}(X) \cap \mathcal{L}_1(X)$ and $\tilde{u} \geq 0$. Moreover, \tilde{u} achieves at least as much revenue as u , and thus it suffices in the above program to look only at feasible u with $u(x^{\text{low}}) = 0$.

We claim that we can therefore remove the constraint $u \geq 0$ and equivalently focus on solving

$$\sup_{u \in \mathcal{U}(X) \cap \mathcal{L}_1(X)} \int_X [\nabla u(x) \cdot x - (u(x) - u(x^{\text{low}}))] f(x) dx. \quad (3.1)$$

Indeed, this objective function agrees with the prior one whenever $u(x^{\text{low}}) = 0$. Furthermore, for any $u \in \mathcal{U}(X) \cap \mathcal{L}_1(X)$, the function $\tilde{u}(x) = u(x) - u(x^{\text{low}})$ is nonnegative and achieves the same objective value. Applying the divergence theorem as in [MV06] we may rewrite the expression for expected revenue in (3.1) as follows:

$$\begin{aligned} \int_X [\nabla u(x) \cdot x - (u(x) - u(x^{\text{low}}))] f(x) dx = & \quad (3.2) \\ \int_{\partial X} u(x) f(x) (x \cdot \hat{n}) dx - \int_X u(x) (\nabla f(x) \cdot x + (n+1)f(x)) dx + u(x^{\text{low}}) \end{aligned}$$

where \hat{n} denotes the outer unit normal field to the boundary ∂X . To simplify notation we make the following definition.

Definition 3 (Transformed measure). *The transformed measure of f is the (signed) measure μ (supported within X) given by the property that*

$$\mu(A) \triangleq \int_{\partial X} \mathbb{I}_A(x) f(x) (x \cdot \hat{n}) dx - \int_X \mathbb{I}_A(x) (\nabla f(x) \cdot x + (n+1)f(x)) dx + \mathbb{I}_A(x^{\text{low}}) \quad (3.3)$$

for all measurable sets A .²

Interpretation of Transformed Measure: Given (3.2) and (3.3), the revenue of the seller in Formulation (3.1) can be written as $\int_X u d\mu$, which is a linear functional of u with respect to the measure μ . Hence, we will maintain the following intuition of what measure μ represents:

“Measure μ quantifies the marginal change in revenue with respect to marginal changes in the rent paid to subsets of buyer types.”

Moreover, our measure satisfies that $\mu(X) = \int_X 1 d\mu = 0$. Indeed, if we substitute $u(x) = 1$ to the left hand side of (3.2), we have that

$$\int_X [\nabla u(x) \cdot x - (u(x) - u(x^{\text{low}}))] f(x) dx = 0.$$

Furthermore, we have $|\mu|(X) < \infty$, since f , ∇f and X are bounded.

Summarizing the above derivation, we obtain the following theorem.

Theorem 1 (Multi-Item Monopoly Problem). *The problem of determining the optimal IC and IR mechanism for a single additive buyer whose values for n goods are distributed according to the joint distribution $f : X \rightarrow \mathbb{R}_{\geq 0}$ is equivalent to solving the optimization problem*

²It follows from boundedness of f 's partial derivatives that μ is a Radon measure. All “measures” we use will be Radon measures.

$$\sup_{u \in \mathcal{U}(X) \cap \mathcal{L}_1(X)} \int_X u d\mu \quad (3.4)$$

where μ is the transformed measure of f given in (3.3).

3.1.3 Example

Consider n independently distributed items, where the value of each item i is drawn uniformly from the bounded interval $[a_i, b_i]$ with $0 \leq a_i < b_i < \infty$. The support of the joint distribution is the set $X = \prod_i [a_i, b_i]$.

For notational convenience, define $v \triangleq \prod_i (b_i - a_i)$, the volume of X . The joint distribution of the items is given by the constant density function f taking value $1/v$ throughout X . The transformed measure μ of f is given by the relation

$$\mu(A) = \mathbb{I}_A(a_1, \dots, a_n) + \frac{1}{v} \int_{\partial X} \mathbb{I}_A(x) (x \cdot \hat{n}) dx - \frac{n+1}{v} \int_X \mathbb{I}_A(x) dx$$

for all measurable sets A . Therefore, by Theorem 1, the optimal revenue is equal to $\sup_{u \in \mathcal{U}(X) \cap \mathcal{L}_1(X)} \int_X u d\mu$, where μ is the sum of:

- A point mass of $+1$ at the point (a_1, \dots, a_n) .
- A mass of $-(n+1)$ distributed uniformly throughout the region X .
- A mass of $+\frac{b_i}{b_i - a_i}$ distributed uniformly on each surface $\{x \in \partial X : x_i = b_i\}$.
- A mass of $-\frac{a_i}{b_i - a_i}$ distributed uniformly on each surface $\{x \in \partial X : x_i = a_i\}$.

Chapter 4

A Dual to Mechanism Design

4.1 A Mechanism Design Strong Duality Theorem

Thus far, we have compactly formulated the problem facing the multi-item monopolist as an optimization problem with respect to the buyer’s utility function; see Formulation (3.4). Unfortunately, the problem is infinite dimensional and cannot be solved directly. Moreover, the problem is not strictly convex so we cannot characterize its optimum using first order conditions. This is an important point of departure in comparison with the work of Rochet and Choné [RC98], where the strict convexity of the cost function allowed first order conditions to drive the characterization. For more discussion see Section 2.2.

In the absence of strict convexity, our approach is to use duality theory. We are seeking to identify a minimization problem, called “the dual problem,” and which is linked to Formulation (3.4), henceforth called “the primal problem,” as follows:

1. We want that the value of any solution to the dual problem is larger than the revenue achieved by any solution to the primal problem. If a minimization problem satisfies this property, it is called a “weak dual problem.”
2. Additionally, we want that the optimum of the dual problem matches the optimum of the primal problem. A minimization problem satisfying this property is called a “strong dual problem.” It is clear that a strong dual problem is also

a weak dual problem. This type of strong dual problem is what we will identify in Theorem 2 of this section.

The importance of identifying a strong dual problem is the following. Given a candidate optimal mechanism, we are guaranteed that a solution to the dual problem with a matching objective value exists if and only if the candidate mechanism is indeed optimal. Therefore, solutions to the dual problem constitute “certificates of optimality” for solutions to the primal, and strong duality guarantees that such dual certificates are always possible to find for optimal solutions to the primal. Accordingly, we will be seeking solutions to our dual problem from Theorem 2 to obtain “certificates, or witnesses, of optimality” for candidate optimal mechanisms. By this we mean that we will be seeking solutions to the dual that prove (via duality theory) that a candidate optimal mechanism is indeed optimal. Moreover, these dual solutions take the form of optimal transportation maps between submeasures induced by measure μ of Definition 3. This tight connection between optimal mechanisms (primal solutions) and optimal transportation maps (dual solutions) drives our characterization of optimal mechanisms in Theorem 3, as well as the concrete examples we work out in Sections 4.3, 5.2 and 6.2. Moreover, by “reverse-engineering the duality theorem” we provide a framework for identifying optimal mechanisms in Section 6.1.

Recent work has applied duality theory to identify optimal mechanisms in the same setting as ours [MV06, DDT13, GK14], albeit this work is restricted in that they only provide weak dual problems. These approaches remove constraints related to truthfulness from the primal formulation, and identify weak dual formulations to such relaxed primal formulations. As such, they provide no guarantee that they can identify dual certificates of optimality for optimal mechanisms. Indeed, while these techniques suffice in certain settings (namely when the constraints removed from the primal happen not to be binding at the optimum), there are simple examples where they fail to apply. Section 4.3.2 provides such a two-item example with uniformly distributed values. In contrast to prior work, we achieve strong duality for the (unrelaxed) primal formulation and our approach is always guaranteed to work.

In this section, we show how to pin down the right dual formulation for the problem

and prove strong duality. The proof of the result requires many analytical tools from measure theory. We give a rough sketch of the proof in this section and postpone the more technical details to Section 4.4.

4.1.1 Measure-Theoretic Preliminaries

We start with some useful measure-theoretic notation:

- $\Gamma(X)$ and $\Gamma_+(X)$ denote the sets of signed and unsigned (Radon) measures on X .
- Given an unsigned measure $\gamma \in \Gamma_+(X \times X)$, we denote by γ_1, γ_2 the two marginals of γ , i.e. $\gamma_1(A) = \gamma(A \times X)$ and $\gamma_2(A) = \gamma(X \times A)$ for all measurable sets $A \subseteq X$.
- For a (signed) measure μ and a measurable $A \subseteq X$, we define the *restriction of μ to A* , denoted $\mu|_A$, by the property $\mu|_A(S) = \mu(A \cap S)$ for all measurable S .
- For a signed measure μ , we will denote by μ_+, μ_- the positive and negative parts of μ , respectively. That is, $\mu = \mu_+ - \mu_-$, where μ_+ and μ_- provide mass to disjoint subsets of X .

We will also be needing certain stochastic dominance properties, namely first- and second-order stochastic dominance as well as the notion of convex dominance.

Definition 4. We say that α first-order (respectively second-order) dominates β for $\alpha, \beta \in \Gamma(X)$, denoted $\alpha \succeq_1 \beta$ (respectively $\alpha \succeq_2 \beta$), if for all non-decreasing continuous (respectively non-decreasing concave) functions $u : X \rightarrow \mathbb{R}$, $\int u d\alpha \geq \int u d\beta$.

Similarly, for vector random variables A and B with values in X , we say that $A \succeq_1 B$ (respectively $A \succeq_2 B$) if $\mathbb{E}[u(A)] \geq \mathbb{E}[u(B)]$ for all non-decreasing continuous (respectively non-decreasing concave) functions $u : X \rightarrow \mathbb{R}$.

Definition 5. We say that α convexly dominates β for $\alpha, \beta \in \Gamma(X)$, denoted $\alpha \succeq_{cvx} \beta$, if for all (non-decreasing, convex) functions $u \in \mathcal{U}(X)$, $\int u d\alpha \geq \int u d\beta$.

Similarly, for vector random variables A and B with values in X , we say that $A \succeq_{cvx} B$ if $\mathbb{E}[u(A)] \geq \mathbb{E}[u(B)]$ for all $u \in \mathcal{U}(X)$.

Interpretation of Convex Dominance: For intuition, a measure $\alpha \succeq_{cvx} \beta$ if we can transform β to α by doing the following two operations:

1. sending (positive) mass to coordinatewise larger points: this makes the integral $\int u d\beta$ larger since u is non-decreasing.
2. spreading (positive) mass so that the mean is preserved: this makes the integral $\int u d\beta$ larger since u is convex.

The existence of a valid transformation using the above operations is equivalent to convex dominance. This follows by Strassen's theorem (presented in Lemma 5).

4.1.2 Mechanism Design Duality

The main result of this work is that the mechanism design problem, formulated as a maximization problem in Theorem 1, has a strong dual problem, as follows:

Theorem 2 (Strong Duality Theorem). *Let $\mu \in \Gamma(X)$ be the transformed measure of the probability density f according to Definition 3. Then*

$$\sup_{u \in \mathcal{U}(X) \cap \mathcal{L}_1(X)} \int_X u d\mu = \inf_{\substack{\gamma \in \Gamma_+(X \times X) \\ \gamma_1 - \gamma_2 \succeq_{cvx} \mu}} \int_{X \times X} \|x - y\|_1 d\gamma(x, y) \quad (4.1)$$

and both the supremum and infimum are achieved. Moreover, the infimum is achieved for some γ^* such that $\gamma_1^*(X) = \gamma_2^*(X) = \mu_+(X)$, $\gamma_1^* \succeq_{cvx} \mu_+$, and $\gamma_2^* \preceq_{cvx} \mu_-$.

Interpretation of the Strong Dual Problem: The dual problem of minimizing $\int \|x - y\|_1 d\gamma$ is an optimization problem that can be intuitively thought as a two step process:

Step 1: Transform μ into a new measure μ' with $\mu'(X) = 0$ such that $\mu' \succeq_{cvx} \mu$. This step is similar to sweeping as defined in [RC98] where they transform the original measure by mean-preserving spreads. However, here we are also allowed to perform positive mass transfers to coordinatewise larger points.

Step 2: Find a joint measure $\gamma \in \Gamma_+(X \times X)$ with $\gamma_1 = \mu'_+$, $\gamma_2 = \mu'_-$ such that

$\int \|x - y\|_1 d\gamma(x, y)$ is minimized. This is an optimal mass transportation problem where the cost of transporting a unit of mass from a point x to a point y is the ℓ_1 distance $\|x - y\|_1$, and we are asked for the cheapest method of transforming the positive part of μ' into the negative part of μ' . Transportation problems of this form have been studied in the mathematical literature. See [Vil08].

Overall, our goal in the dual problem is to match the positive part of μ to the negative part of μ at a minimum cost where some operations come for free, namely we can choose any $\mu' \succeq_{cvx} \mu$ that is convenient to us, foreseeing that transporting μ'_+ to μ'_- comes at a cost equal to the total ℓ_1 distance that mass travels.

We remark that establishing that the right hand side of (4.1) is a weak dual for the left hand side is easy. Proving strong duality is significantly more challenging, and relies on non-trivial analytical tools such as the Fenchel-Rockafellar duality theorem. We postpone the proof to Section 4.4, and proceed to show weak duality.

Lemma 1 (Weak Duality). *Let $\mu \in \Gamma(X)$. Then*

$$\sup_{u \in \mathcal{U}(X) \cap \mathcal{L}_1(X)} \int_X u d\mu \leq \inf_{\substack{\gamma \in \Gamma_+(X \times X) \\ \gamma_1 - \gamma_2 \succeq_{cvx} \mu}} \int_{X \times X} \|x - y\|_1 d\gamma.$$

PROOF OF LEMMA 1: For any feasible u for the left-hand side and feasible γ for the right-hand side, we have

$$\int_X u d\mu \leq \int_X u d(\gamma_1 - \gamma_2) = \int_{X \times X} (u(x) - u(y)) d\gamma(x, y) \leq \int_{X \times X} \|x - y\|_1 d\gamma(x, y)$$

where the first inequality follows from $\gamma_1 - \gamma_2 \succeq_{cvx} \mu$ and the second inequality follows from the 1-Lipschitz condition on u . \square

From the proof of Lemma 1, we note the following “complementary slackness” conditions that a pair of optimal primal and dual solutions must satisfy.

Corollary 1. *Let u^* and γ^* be feasible for their respective problems above. Then $\int u^* d\mu = \int \|x - y\|_1 d\gamma^*$ if and only if both of the following conditions hold:*

1. $\int u^* d(\gamma_1^* - \gamma_2^*) = \int u^* d\mu.$
2. $u^*(x) - u^*(y) = \|x - y\|_1, \gamma^*(x, y)$ -almost surely.

PROOF OF COROLLARY 1: The inequalities in the proof of Lemma 1 are tight precisely when both conditions hold. \square

Interpretation of the Complementary Slackness Conditions

Remark 1. It is useful to geometrically interpret Corollary 1:

Condition 1: We view $\gamma_1^* - \gamma_2^*$ (denote this by μ') as a “shuffled” μ . Stemming from the $\mu' \succeq_{cvx} \mu$ constraint, the shuffling of μ into μ' is obtained via any sequence of the following operations: (1) Picking a positive point mass δ_x from μ_+ and sending it from point x to some other point $y \geq x$ (coordinate-wise). The constraint $\int u^* d\mu' = \int u^* d\mu$ requires that $u^*(x) = u^*(y)$. Recall that u^* is non-decreasing, so $u^*(z) = u^*(x)$ for all $z \in \prod_j [x_j, y_j]$. Thus, if y is strictly larger than x in coordinate i , then $(\nabla u^*)_i = 0$ at all points z “in between” x and y . The other operation we are allowed, called a “mean-preserving spread,” is (2) picking a positive point mass δ_x from μ_+ , splitting the point mass into several pieces, and sending these pieces to multiple points while preserving the center of mass. The constraint $\int u^* d\mu' = \int u^* d\mu$ requires that u^* varies *linearly* between x and all points z that received a piece.

Condition 2: The second condition is more straightforward than the first. We view γ^* as a “transport” map between its component measures γ_1^* and γ_2^* . The condition states that if γ^* transports from location x to location y , then $u^*(x) = u^*(y) + \|x - y\|_1$. If for some coordinate $i, x_i < y_i$, then $\|z - y\|_1 < \|x - y\|_1$ for z with $z_j = \max(x_j, y_j)$. This leads to a contradiction since $u^*(x) - u^*(y) \leq u^*(z) - u^*(y) \leq \|z - y\|_1 < \|x - y\|_1$. Therefore, it must be the case that (1) x is component-wise greater than or equal to y and (2) if $x_i > y_i$ in coordinate i , then $(\nabla u^*)_i = 1$ at all points “in between” x and y . That is, the mechanism allocates item i with probability 1 to all those types.

By Lemma 1 and Corollary 1, if we can find a “tight pair” of u^* and γ^* , then they are optimal for their respective problems. This is useful since constructing a γ

that satisfies the conditions of Corollary 1 serves as a certificate of optimality for a mechanism. Theorem 2 shows that this approach always works: for any optimal u^* there always exists a γ^* satisfying the conditions of Corollary 1.

Remark 2. It is useful to discuss what in our dual formulation in the RHS of (4.1) makes it a strong dual, comparing to the previous work [DDT13, GK14]. If we were to tighten the $\gamma_1 - \gamma_2 \succeq_{cvx} \mu$ constraint in our dual formulation to a first-order stochastic dominance constraint, we essentially recover the duality framework of [DDT13, GK14]. Tightening the dual constraint, maintains the weak duality but creates a gap between the optimal primal and dual values. In particular, the dual problem resulting from tightening this constraint becomes a strong dual problem for a relaxed version of the mechanism design problem in which the convexity constraint on u is dropped.

4.2 Single-Item Applications and Interpretation

Before considering multi-item settings, it is instructive to study the application of our strong duality theorem to single-item settings. We seek to relate the task of minimizing the transportation cost in the dual problem from Theorem 2 to the structure of Myerson's solution [Mye81].

Consider the task of selling a single item to a buyer whose value z for the item is distributed according to a twice-differentiable regular distribution F supported on $[\underline{z}, \bar{z}]$.¹ Since $n = 1$, if we were to apply our duality framework to this setting, we would choose μ according to (3.3) as follows:

$$\begin{aligned} \mu(A) &= \mathbb{I}_A(\underline{z}) \cdot (1 - f(\underline{z}) \cdot \underline{z}) + \mathbb{I}_A(\bar{z}) \cdot f(\bar{z}) \cdot \bar{z} - \int_{\underline{z}}^{\bar{z}} \mathbb{I}_A(z) (f'(z) \cdot z + 2f(z)) dz \\ &= \mathbb{I}_A(\underline{z}) \cdot (1 - f(\underline{z}) \cdot \underline{z}) + \mathbb{I}_A(\bar{z}) \cdot f(\bar{z}) \cdot \bar{z} - \int_{\underline{z}}^{\bar{z}} \mathbb{I}_A(z) \left(\left(z - \frac{1 - F(z)}{f(z)} \right) f'(z) \right)' dz \end{aligned}$$

¹We remind the reader that a differentiable distribution F is *regular* when its Myerson virtual value function $\phi(z) = z - \frac{1 - F(z)}{f(z)}$ is increasing in its support, where f is the distribution density function.

We can interpret the transportation problem of Theorem 2, defined in terms of μ , as:

- The sub-population of buyers having the right-most type, \bar{z} , in the support of the distribution have an excess supply of $f(\bar{z}) \cdot \bar{z}$;
- The sub-population of buyers with the left-most type, \underline{z} , in the support have an excess supply of $1 - f(\underline{z}) \cdot \underline{z}$;
- Finally, the sub-population of buyers at each other type, z , have a demand of

$$\left(\left(z - \frac{1 - F(z)}{f(z)} \right) f(z) \right)' dz$$

One way to satisfy the above supply/demand requirements is to have every infinitesimal buyer of type z push mass of $z - \frac{1 - F(z)}{f(z)}$ to its left. Since the fraction of buyers at z is $f(z)$, the total amount of mass staying with them is then $\left(\left(z - \frac{1 - F(z)}{f(z)} \right) f(z) \right)' dz$ as required. Notice, in particular, that buyers with positive virtual types will push mass to their left, while buyers with negative virtual types will push mass to their right.

The afore-described transportation map is feasible for our transportation problem as it satisfies all demand/supply constraints. We also claim that this solution is optimal. To see this consider the mechanism that allocates the item to all buyers with non-negative virtual type at a fixed price p^* . The resulting utility function is of the form $\max\{z - p^*, 0\}$. We claim that this utility function satisfies the complementary slackness conditions of Remark 1 with respect to the transportation map identified above. Indeed, when $z > p^*$, u is linear with $u'(z) = 1$ and mass is sent to the left—which is allowed by Part 2 of the remark, while, when $z < p^*$, u is 0 with $u'(z) = 0$ and mass is sent to the right—allowed by Part 1(1) of the remark.

In conclusion, when F is regular, the virtual values dictate exactly how to optimally solve the optimal transportation problem of Theorem 2. Each infinitesimal buyer of type z will push mass that equals its virtual value to its left. In particular, the optimal transportation does not need to use mean-preserving spreads. Moreover, measure μ can be interpreted as the “negative marginal normalized virtual value,” as

it assigns measure $-\left(\left(z - \frac{1-F(z)}{f(z)}\right) f(z)\right)' dz$ to the interval $[z, z + dz]$, when $z \neq \underline{z}, \bar{z}$.

When F is not regular, the afore-described transportation map is not optimal due to the non-monotonicity of the virtual values. In this case, we need to pre-process our measure μ via mean-preserving spreads, prior to the transport, and ironing dictates how to do these mean-preserving spreads. In other words, ironing dictates how to perform the sweeping of the type set prior to transport.

4.3 Multi-Item Applications of Duality

We now give two examples of using Theorem 2 to prove optimality of mechanisms for selling two uniformly distributed independent items.

4.3.1 Two Uniform $[0, 1]$ Items

Using Theorem 2, we provide a short proof of optimality of the mechanism for two i.i.d. uniform $[0, 1]$ items proposed by [MV06] which we refer to as the MV-mechanism:

Example 1. *The optimal IC and IR mechanism for selling two items whose values are distributed uniformly and independently on the interval $[0, 1]$ is the following menu:*

- *buy any single item for a price of $\frac{2}{3}$; or*
- *buy both items for a price of $\frac{4-\sqrt{2}}{3}$.*

Let Z be the set of types that receive no goods and pay 0 to the MV-mechanism. Also, let A, B be the set of types that receive only goods 1 and 2 respectively and W be the set of types that receive both goods. The sets A, B, Z, W are illustrated in Figure 4-1 and separated by solid lines.

Let us now try to prove that the MV mechanism is indeed optimal. As a first step, we need to compute the transformed measure μ of the uniform distribution on $[0, 1]^2$. We have already computed μ in Section 3.1.3. It has a point mass of +1 at $(0, 0)$, a mass of -3 distributed uniformly over $[0, 1]^2$, a mass of +1 distributed uniformly on the top boundary of $[0, 1]^2$, and a mass of +1 distributed uniformly on the right

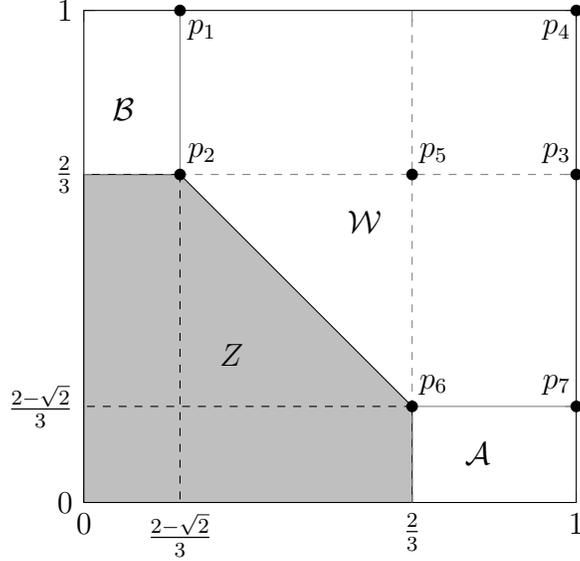


Figure 4-1: The MV-mechanism for two i.i.d. uniform $[0, 1]$ items.

boundary. Notice that the total net mass is equal to 0 within each region Z , A , B , or W .

To prove optimality of the MV-mechanism, we will construct an optimal γ^* for the dual program of Theorem 2 to match the positive mass μ_+ to the negative μ_- . Our γ^* will be decomposed into $\gamma^* = \gamma^Z + \gamma^A + \gamma^B + \gamma^W$ and to ensure that $\gamma_1^* - \gamma_2^* \succeq_{cvx} \mu$, we will show that

$$\gamma_1^Z - \gamma_2^Z \succeq_{cvx} \mu|_Z; \quad \gamma_1^A - \gamma_2^A \succeq_{cvx} \mu|_A; \quad \gamma_1^B - \gamma_2^B \succeq_{cvx} \mu|_B; \quad \gamma_1^W - \gamma_2^W \succeq_{cvx} \mu|_W.$$

We will also show that the conditions of Corollary 1 hold for each of the measures γ^Z , γ^A , γ^B , and γ^W separately, namely $\int u^* d(\gamma_1^S - \gamma_2^S) = \int_S u^* d\mu$ and $u^*(x) - u^*(y) = \|x - y\|_1$ hold γ^S -almost surely for $S = Z, A, B$, and W .

Construction of γ^Z : Since $\mu_+|_Z$ is a point-mass at $(0, 0)$ and $\mu_-|_Z$ is distributed throughout a region which is coordinatewise greater than $(0, 0)$, we notice that $\mu|_Z \preceq_{cvx} 0$. We set γ^Z to be the zero measure, and the relation $\gamma_1^Z - \gamma_2^Z = 0 \succeq_{cvx} \mu|_Z$, as well as the two necessary equalities from Corollary 1, are trivially satisfied.

Construction of γ^A and γ^B : In region A , $\mu_+|_A$ is distributed on the right boundary

while $\mu_-|_A$ is distributed uniformly on the interior of A . We construct γ^A by transporting the positive mass $\mu_+|_A$ to the left to match the negative mass $\mu_-|_A$. Notice that this indeed matches completely the positive mass to the negative since $\mu(A) = 0$ and intuitively minimizes the ℓ_1 transportation distance. To see that the two necessary equalities from Corollary 1 are satisfied, notice that $\gamma_1^A = \mu_+|_A, \gamma_2^A = \mu_-|_A$ so the first equality holds. The second inequality holds as we are transporting mass only to the left and thus the measure γ^A is concentrated on pairs $(x, y) \in A \times A$ such that $1 = x_1 \geq y_1 \geq \frac{2}{3}$ and $x_2 = y_2$. Moreover, for all such pairs (x, y) , we have that $u(x) - u(y) = (x_1 - \frac{2}{3}) - (y_1 - \frac{2}{3}) = x_1 - y_1 = \|x - y\|_1$. The construction of γ^B is similar.

Construction of γ^W We construct an explicit matching that only matches leftwards and downwards without doing any prior mass shuffling. We match the positive mass on the segment p_1p_4 to the negative mass on the rectangle $p_1p_2p_3p_4$ by moving mass downwards. We match the positive mass of the segment p_3p_7 to the negative mass on the rectangle $p_3p_5p_6p_7$ by moving mass leftwards. Finally, we match the positive mass on the segment p_3p_4 to the negative mass on the triangle $p_2p_5p_6$ by moving mass downwards and leftwards. Notice that all positive/negative mass in region W has been accounted for, all of $(\mu|_W)_+$ has been matched to all of $(\mu|_W)_-$ and all moves were down and to the left, establishing $u(x) - u(y) = (x_1 + x_2 - \frac{4-\sqrt{2}}{3}) - (y_1 + y_2 - \frac{4-\sqrt{2}}{3}) = x_1 + x_2 - y_1 - y_2 = \|x - y\|_1$.

4.3.2 Two Uniform But Not Identical Items

We now present an example with two items whose values are distributed uniformly and independently on the intervals $[4, 16]$ and $[4, 7]$. We note that the distributions are not identical, and thus the characterization of [Pav11] does not apply. In addition, the relaxation-based duality framework of [DDT13, GK14] (see Remark 2) fails in this example: if we were to relax the constraint that the utility function u be convex, the “mechanism design program” would have a solution with greater revenue than is actually possible.

Example 2. *The optimal IC and IR mechanism for selling two items whose values are distributed uniformly and independently on the intervals $[4, 16]$ and $[4, 7]$ is as follows:*

- *If the buyer's declared type is in region Z , he receives no goods and pays nothing.*
- *If the buyer's declared type is in region Y , he pays a price of 8 and receives the first good with probability 50% and the second good with probability 1.*
- *If the buyer's declared type is in region W , he gets both goods for a price of 12.*

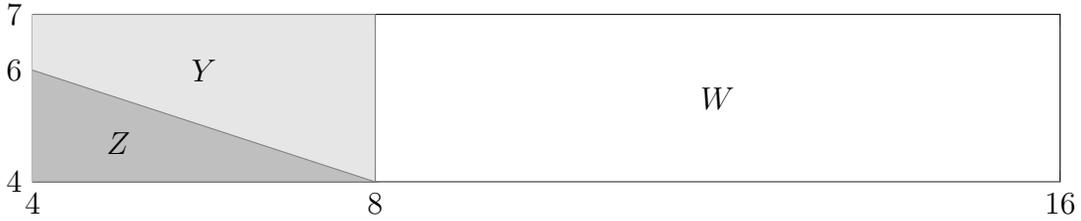


Figure 4-2: Partition of $[4, 16] \times [4, 7]$ into different regions by the optimal mechanism.

The proof of optimality of our proposed mechanism works by constructing a measure $\gamma = \gamma^Z + \gamma^Y + \gamma^W$ separately in each region. The constructions of γ^W and γ^Z are similar to the previous example. The construction of γ^Y , however, is a little more intricate as it requires an initial shuffling of the mass before computing the optimal way to transport the resulting mass.

It is straightforward to verify that the mechanism is IC and IR. All that remains is to prove that the utility function u^* induced by the mechanism is optimal.

The transformed measure μ of the type distribution is composed of:

- A point mass of +1 at $(4, 4)$.
- Mass -3 distributed throughout the rectangle (Density $-\frac{1}{12}$)
- Mass $+\frac{7}{3}$ distributed on upper edge of rectangle (Linear density $+\frac{7}{36}$)
- Mass $-\frac{4}{3}$ distributed on lower edge of rectangle (Linear density $-\frac{1}{9}$)

- Mass $+\frac{4}{3}$ distributed on right edge of rectangle (Linear density $+\frac{4}{9}$)
- Mass $-\frac{1}{3}$ distributed on left edge of rectangle (Linear density $-\frac{1}{9}$)

We claim that $\mu(Z) = \mu(Y) = \mu(W) = 0$, which is straightforward to verify.

We will construct an optimal γ^* for the dual program of Theorem 2, using the intuition of Remark 1. Our γ^* will be decomposed into $\gamma^* = \gamma^Z + \gamma^Y + \gamma^W$ with $\gamma^Z \in \Gamma_+(Z \times Z)$, $\gamma^Y \in \Gamma_+(Y \times Y)$, and $\gamma^W \in \Gamma_+(W \times W)$. To ensure that $\gamma_1^* - \gamma_2^* \succeq_{cvx} \mu$, we will show that

$$\gamma_1^Z - \gamma_2^Z \succeq_{cvx} \mu|_Z; \quad \gamma_1^Y - \gamma_2^Y \succeq_{cvx} \mu|_Y; \quad \gamma_1^W - \gamma_2^W \succeq_{cvx} \mu|_W.$$

We will also show that the conditions of Corollary 1 hold for each of the measures γ^Z , γ^Y , and γ^W separately, namely $\int u^* d(\gamma_1^A - \gamma_2^A) = \int_A u^* d\mu$ and $u^*(x) - u^*(y) = \|x - y\|_1$ hold γ^A -almost surely for $A = Z, Y$, and W .

Construction of γ^Z . Since $\mu_+|_Z$ is a point-mass at $(4, 4)$ and $\mu_-|_Z$ is distributed throughout a region which is coordinatewise greater than $(4, 4)$, we notice that $\mu|_Z \preceq_{cvx} 0$. We therefore set γ^Z to be the zero measure, and the relation $\gamma_1^Z - \gamma_2^Z = 0 \succeq_{cvx} \mu|_Z$, as well as the two necessary equalities from Corollary 1, are trivially satisfied.

Construction of γ^W . We will construct $\gamma^W \in \Gamma(\mu_+|_W, \mu_-|_W)$ such that $x \geq y$ component-wise holds $\gamma^W(x, y)$ almost surely. Geometrically, we view this as “transporting” $\mu_+|_W$ into $\mu_-|_W$ by moving mass downwards and leftwards. Indeed, since both items are allocated with probability 1 in W , being able to transport both downwards and leftwards is in line with our interpretation of the second condition of Corollary 1, as explained in Remark 1.²

We notice that $\mu_+|_W$ consists of mass distributed on the top and right edges of W , while $\mu_-|_W$ consists of mass on the interior and bottom of W . We first match the μ_+ mass on $[8, 16] \times \{7\}$ with the μ_- mass on $[8, 16] \times [\frac{14}{3}, 7]$ by moving mass downwards,

²To prove the existence of such a map, it is equivalent by Strassen’s theorem to prove that $\mu_+|_W$ stochastically dominates $\mu_-|_W$ in the first order, but in this example we will directly define such a map.

then we match the μ_+ mass on $\{16\} \times [4, \frac{14}{3}]$ with the μ_- mass on $[\frac{32}{3}, 16] \times (4, \frac{14}{3}]$ by moving mass to the left, and we finally match the μ_+ mass on $\{16\} \times [\frac{14}{3}, 7]$ with the remaining negative mass arbitrarily. Noticing that $u^*(x) = \|x\|_1 - 12$ for all $x \in W$, it is straightforward to verify the desired properties from Corollary 1.

Construction of γ^Y . This is the most involved step of the proof. Since item 2 is allocated with 100% probability in region Y , by Remark 1 we would like to transport the positive mass $\mu_+|_Y$ into $\mu_-|_Y$ by moving mass straight downwards. However, this is impossible without first “shuffling” $\mu|_Y$, due to the negative mass on the left boundary of Y . Therefore, we first “shuffle” the positive part of $\mu|_Y$ (on the top boundary) to push positive mass onto the point $(4, 7)$ (the top-left corner of Y), and only then do we transport the positive part of the shuffled measure into the negative part by sending mass downwards. Since the positive and negative parts of $\mu|_Y$ must be matchable by only sending mass downwards, we know how the post-shuffling measure should look. In particular, on every vertical line in region Y the net post-shuffling mass should be zero.

So rather than constructing γ^Y with $\gamma_1^Y - \gamma_2^Y$ equal to $\mu|_Y$, we will have $\gamma_1^Y - \gamma_2^Y = \mu|_Y + \alpha$, where the “shuffling” measure $\alpha = \alpha_+ - \alpha_- \succeq_{cvx} 0$. As discussed above, we set α to have density function

$$f_\alpha(z_1, z_2) = \mathbb{I}_{z_2=7} \cdot \left(\frac{1}{9} \mathbb{I}_{z_1=4} + \frac{1}{24} \left(z_1 - \frac{20}{3} \right) \right) \cdot \mathbb{I}_{z \in Y}.$$

The measure α is supported on the line $[4, 8] \times \{7\}$ and consists of a point mass of $\frac{1}{9}$ at $(4, 7)$ followed by allocating mass along the 1-dimensional upper boundary of Y according to a density function which begins negative and increases linearly. It is straightforward to verify that $\alpha \succeq_{cvx} 0$,³ which we need for feasibility, and that

³Since α is supported on a 1-dimensional line, this verification uses a property analogous to the standard characterization of one-dimensional second-order stochastic dominance via the cumulative density function. Informally, we can argue that $\alpha \succeq_{cvx} 0$ by considering integrals of one-dimensional test functions (by restricting our attention to the line $z_2 = 7$) and noticing that, since $\alpha(Y) = 0$, we need only consider test functions h which have value 0 at $z_1 = 4$. We then use the fact that all linear functions integrate to 0 under α and that (ignoring the point mass at $z_1 = 4$, since h is 0 at this point) the density of α is monotonically increasing.

$\int_Y u^* d\alpha = 0$, which we need to satisfy complementary slackness.

We are now ready to define $\gamma^Y \in \Gamma(\mu_+|_Y + \alpha_+, \mu_-|_Y + \alpha_-)$. We construct γ^Y so that $x_1 = y_1$ and $x_2 \geq y_2$ hold $\gamma^Y(x, y)$ almost surely. Since $\mu_+|_Y + \alpha_+$ only assigns mass to the upper boundary of Y , to show that γ^Y can be constructed so that all mass is transported “vertically downwards” we need only verify that $\mu_+|_Y + \alpha_+$ and $\mu_-|_Y + \alpha_-$ assign the same density to any vertical “strip” in Y . Indeed,

$$\begin{aligned} (\mu_-|_Y + \alpha_-)(\{4\} \times [6, 7]) &= \mu_-|_Y(\{4\} \times [6, 7]) = \frac{1}{9} = \alpha_+(\{4\} \times [6, 7]) \\ &= (\mu_+|_Y + \alpha_+)(\{4\} \times [6, 7]) \end{aligned}$$

and, for all $z_1 \pm \epsilon \in (4, 8]$, we compute the following, using the fact that the surface area of $Y \cap ([z_1 - \epsilon, z_1 + \epsilon] \times [4, 7])$ is $2\epsilon \cdot (\frac{z_1}{2} - 1)$:

$$\begin{aligned} &(\mu_-|_Y + \alpha_-)([z_1 - \epsilon, z_1 + \epsilon] \times [4, 7]) \\ &= \frac{1}{12} \cdot \left(2\epsilon \cdot \left(\frac{z_1}{2} - 1\right)\right) - \frac{1}{24} \int_{z_1 - \epsilon}^{z_1 + \epsilon} \left(z - \frac{20}{3}\right) dz \\ &= \frac{\epsilon z_1}{12} - \frac{\epsilon}{6} - \frac{1}{24} (2\epsilon z_1 - \frac{40\epsilon}{3}) = \frac{7\epsilon}{18} = \mu_+|_Y([z_1 - \epsilon, z_1 + \epsilon] \times [4, 7]). \end{aligned}$$

Since u^* has the property that $u^*(z_1, a) - u^*(z_1, b) = a - b$ for all $(z_1, a), (z_1, b) \in Y$ (as the second good is received with probability 1), it follows that γ^Y satisfies the necessary conditions of Corollary 1.

4.3.3 Discussion

Our examples in this section serve to illustrate how to use our duality theorem to verify the optimality of our proposed mechanisms, without explaining how we identified these mechanisms. These mechanisms were in fact identified by “reverse-engineering” the duality theorem. The next two chapters provide tools for performing this reverse-engineering. In particular, Section 5.1 provides a characterization of mechanism optimality in terms of stochastic dominance conditions satisfied in regions partitioning the type space. Alleviating the need to reverse-engineer the duality theorem, Sec-

tion 6.1 prescribes a straightforward procedure for identifying optimal mechanisms. We use this procedure to solve several examples in Section 6.2.

4.4 Proof of Strong Mechanism Design Duality

In this section, we give a formal proof of the strong mechanism duality theorem, Theorem 2. To carefully prove the statement, we specify that the proof is for Radon measures. A Radon measure is a locally-finite inner-regular Borel measure. We use $\Gamma(X) = \text{Radon}(X)$ (resp. $\Gamma_+(X) = \text{Radon}_+(X)$) as the set of signed (resp. unsigned) Radon measures on X . The transformed measure of a distribution is always a signed Radon measure as it defines a bounded linear functional on the utility function u .⁴

4.4.1 A Strong Duality Lemma

The overall structure of our proof of Theorem 2 is roughly parallel to the proof of Monge-Kantorovich duality presented in [Vil08], although the technical aspects of our proof are different, mainly due to the added convexity constraint on u . We begin by stating the Legendre-Fenchel transformation and the Fenchel-Rockafellar duality theorem.

Definition 6 (Legendre-Fenchel Transform). *Let E be a normed vector space and let $\Lambda : E \rightarrow \mathbb{R} \cup \{+\infty\}$ be a convex function. The Legendre-Fenchel transform of Λ , denoted Λ^* , is a map from the topological dual E^* of E to $\mathbb{R} \cup \{\infty\}$ given by*

$$\Lambda^*(z^*) = \sup_{z \in E} (\langle z^*, z \rangle - \Lambda(z)).$$

Claim 1 (Fenchel-Rockafellar duality). *Let E be a normed vector space, E^* its topological dual, and Θ, Ξ two convex functions on E taking values in $\mathbb{R} \cup \{+\infty\}$. Let Θ^*, Ξ^* be the Legendre-Fenchel transforms of Θ and Ξ respectively. Assume that there*

⁴More formally, this follows from Riesz representation theorem

exists $z_0 \in E$ such that $\Theta(z_0) < +\infty$, $\Xi(z_0) < +\infty$ and Θ is continuous at z_0 . Then

$$\inf_{z \in E} [\Theta(z) + \Xi(z)] = \max_{z^* \in E^*} [-\Theta^*(-z^*) - \Xi^*(z^*)].$$

Lemma 2. *Let X be a compact convex subset of \mathbb{R}^n , and let $\mu \in \Gamma(X)$ be such that $\mu(X) = 0$. Then*

$$\inf_{\substack{\gamma \in \Gamma_+(X \times X) \\ \gamma_1 \succeq_{cvx} \mu_+ \\ \gamma_2 \preceq_{cvx} \mu_-}} \int_{X \times X} \|x - y\|_1 d\gamma(x, y) = \sup_{\substack{\phi, \psi \in \mathcal{U}(X) \\ \phi(x) - \psi(y) \leq \|x - y\|_1}} \left(\int_X \phi d\mu_+ - \int_X \psi d\mu_- \right)$$

and the infimum on the left-hand side is achieved.

PROOF OF LEMMA 2:

We will apply Fenchel-Rockafellar duality with $E = CB(X \times X)$, the space of continuous (and bounded) functions on $X \times X$ equipped with the $\|\cdot\|_\infty$ norm. Since X is compact, by the Riesz representation theorem $E^* = \Gamma(X \times X)$.

We now define functions Θ, Ξ mapping $CB(X \times X)$ to $\mathbb{R} \cup \{+\infty\}$ by

$$\Theta(f) = \begin{cases} 0 & \text{if } f(x, y) \geq -\|x - y\|_1 \text{ for all } x, y \in X \\ +\infty & \text{otherwise} \end{cases}$$

$$\Xi(f) = \begin{cases} \int_X \psi d\mu_- - \int_X \phi d\mu_+ & \text{if } f(x, y) = \psi(y) - \phi(x) \text{ for some } \psi, \phi \in \mathcal{U}(X) \\ +\infty & \text{otherwise.} \end{cases}$$

We note that Ξ is well-defined: If $\psi(x) - \phi(y) = \psi'(x) - \phi'(y)$ for all $x, y \in X$, then $\psi(x) - \psi'(x) = \phi(y) - \phi'(y)$ for all $x, y \in X$. This means that ψ' differs from ψ only by an additive constant, and ϕ differs from ϕ' by the same additive constant, and therefore (since μ_+ and μ_- have the same total mass) $\int_X \psi d\mu_- - \int_X \phi d\mu_+ = \int_X \psi' d\mu_- - \int_X \phi' d\mu_+$.

It is clear that $\Theta(f)$ is convex, since any convex combination two functions for which $f(x, y) \geq -\|x - y\|_1$ will yield another function for which the inequality is satisfied. It is furthermore clear that Ξ is convex, since we can take convex combinations of the ψ and ϕ functions as appropriate. (Notice that $\mathcal{U}(X)$ is closed under addition

and positive scaling of functions.)

Consider the function $z_0 \in CB(X \times X)$ which takes the constant value of 1. It is clear that $\Theta(z_0) = 0$ and $\Xi(z_0) = \mu_-(X) < \infty$. Furthermore, $\Theta(z) = 0$ for any $z \in CB(X \times X)$ with $\|z - z_0\|_\infty < 1$, and therefore Θ is continuous at z_0 . We can thus apply the Fenchel-Rockafellar duality theorem.

We compute, for any $\gamma \in \Gamma(X \times X)$:

$$\begin{aligned} \Theta^*(-\gamma) &= \sup_{f \in CB(X \times X)} \left[\int_{X \times X} f(x, y) d(-\gamma(x, y)) \right. \\ &\quad \left. - \begin{cases} 0 & \text{if } f(x, y) \geq -\|x - y\|_1 \ \forall x, y \in X \\ +\infty & \text{otherwise} \end{cases} \right] \\ &= \sup_{\substack{f \in CB(X \times X) \\ f(x, y) \geq -\|x - y\|_1}} \left(- \int_{X \times X} f(x, y) d\gamma(x, y) \right) = \sup_{\substack{\tilde{f} \in CB(X \times X) \\ \tilde{f}(x, y) \leq \|x - y\|_1}} \left(\int_{X \times X} \tilde{f}(x, y) d\gamma(x, y) \right). \end{aligned}$$

We claim therefore that

$$\Theta^*(-\gamma) = \begin{cases} \int_{X \times X} \|x - y\|_1 d\gamma(x, y) & \text{if } \gamma \in \Gamma_+(X \times X) \\ \infty & \text{otherwise.} \end{cases}$$

Indeed, if γ is a positive linear functional, then the result follows from monotonicity, since $\|x - y\|_1$ is the pointwise greatest function \tilde{f} satisfying the constraint $\tilde{f}(x, y) \leq \|x - y\|_1$, and $\|x - y\|_1$ is continuous. Suppose instead that γ is a signed Radon measure which is not positive everywhere. Then there exists a continuous nonnegative function $g : X \times X \rightarrow \mathbb{R}$ such that $\int g d\gamma = -\epsilon$ for some $\epsilon > 0$.⁵ Since $g(x, y) \geq 0$, it follows that $-kg(x, y) \leq 0 \leq \|x - y\|_1$ for any $k \geq 0$. Therefore

$$\sup_{\substack{\tilde{f} \in CB(X \times X) \\ \tilde{f}(x, y) \leq \|x - y\|_1}} \left(\int_{X \times X} \tilde{f}(x, y) d\gamma(x, y) \right) \geq \int -kg(x, y) d\gamma(x, y) = k\epsilon.$$

The claim follows, since $k > 0$ is arbitrary.

⁵Formally, we have used Lusin's theorem to find such a g which is continuous, as opposed to merely measurable.

We similarly compute, for any $\gamma \in \Gamma(X \times X)$:

$$\begin{aligned} \Xi^*(\gamma) &= \sup_{f \in CB(X \times X)} \left[\int_{X \times X} f(x, y) d\gamma(x, y) - \right. \\ &\quad \left. - \begin{cases} \int_X \psi d\mu_- - \int_X \phi d\mu_+ & \text{if } f(x, y) = \psi(y) - \phi(x) \text{ and } \psi, \phi \in \mathcal{U}(X) \\ +\infty & \text{otherwise} \end{cases} \right] \\ &= \sup_{\psi, \phi \in \mathcal{U}(X)} \left[\int_{X \times X} (\psi(y) - \phi(x)) d\gamma(x, y) - \int_X \psi d\mu_- + \int_X \phi d\mu_+ \right] \end{aligned}$$

We notice that $\Xi^*(\gamma) \geq 0$ for all $\gamma \in \Gamma(X \times X)$ by setting $\psi = \phi = 0$ and thus $\Theta^*(-\gamma) + \Xi^*(\gamma) = \infty$ if $\gamma \notin \Gamma_+(X \times X)$. Moreover, when $\gamma \in \Gamma_+(X \times X)$:

$$\begin{aligned} \Xi^*(\gamma) &= \sup_{\psi, \phi \in \mathcal{U}(X)} \left[\int_{X \times X} (\psi(y) - \phi(x)) d\gamma(x, y) - \int_X \psi d\mu_- + \int_X \phi d\mu_+ \right] \\ &= \sup_{\psi, \phi \in \mathcal{U}(X)} \left[\int_X \psi d(\gamma_2 - \mu_-) + \int_X \phi d(\mu_+ - \gamma_1) \right] \\ &= \begin{cases} 0 & \text{if } \gamma_1 \succeq_{cvx} \mu_+ \text{ and } \gamma_2 \preceq_{cvx} \mu_- \\ \infty & \text{otherwise.} \end{cases} \end{aligned}$$

The last equality is true because if $\gamma_1 \succeq_{cvx} \mu_+$ doesn't hold, we can find a function $\phi \in \mathcal{U}(X)$ such that $\int_X \phi d(\mu_+ - \gamma_1) > 0$. Since we are allowed to scale ϕ arbitrarily, we can make the inside quantity as large as we want. The same holds when $\mu_- \not\preceq_{cvx} \gamma_2$.

We now apply Fenchel-Rockafellar duality:

$$\begin{aligned} \inf_{f \in CB(X \times X)} [\Theta(f) + \Xi(f)] &= \max_{\gamma \in \Gamma(X \times X)} [-\Theta^*(-\gamma) - \Xi^*(\gamma)] \\ \inf_{\substack{f(x, y) \geq -\|x-y\|_1 \\ f(x, y) = \psi(y) - \phi(x) \\ \psi, \phi \in \mathcal{U}(X)}} \left(\int_X \psi d\mu_- - \int_X \phi d\mu_+ \right) &= \max_{\gamma \in \Gamma_+(X \times X)} \left[- \int_{X \times X} \|x - y\|_1 d\gamma(x, y) - \Xi^*(\gamma) \right] \\ \inf_{\substack{\psi, \phi \in \mathcal{U}(X) \\ \phi(x) - \psi(y) \leq \|x-y\|_1}} \left(\int_X \psi d\mu_- - \int_X \phi d\mu_+ \right) &= \max_{\substack{\gamma \in \Gamma_+(X \times X) \\ \gamma_1 \succeq_{cvx} \mu_+ \\ \gamma_2 \preceq_{cvx} \mu_-}} \left(- \int_{X \times X} \|x - y\|_1 d\gamma(x, y) \right) \\ \sup_{\substack{\psi, \phi \in \mathcal{U}(X) \\ \phi(x) - \psi(y) \leq \|x-y\|_1}} \left(\int_X \phi d\mu_+ - \int_X \psi d\mu_- \right) &= \min_{\substack{\gamma \in \Gamma_+(X \times X) \\ \gamma_1 \succeq_{cvx} \mu_+ \\ \gamma_2 \preceq_{cvx} \mu_-}} \left(\int_{X \times X} \|x - y\|_1 d\gamma(x, y) \right). \end{aligned}$$

□

4.4.2 From Two Convex Functions to One

Lemma 3. *Let $X = \prod_{i=1}^n [x_i^{low}, x_i^{high}]$ for some $x_i^{low}, x_i^{high} \geq 0$, and let $\mu \in \Gamma(X)$ such that $\mu(X) = 0$. Then*

$$\sup_{\substack{\phi, \psi \in \mathcal{U}(X) \\ \phi(x) - \psi(y) \leq \|x - y\|_1}} \left(\int_X \phi d\mu_+ - \int_X \psi d\mu_- \right) = \sup_{u \in \mathcal{U}(X) \cap \mathcal{L}_1(X)} \left(\int_X u d\mu_+ - \int_X u d\mu_- \right).$$

Furthermore, if the supremum of one side is achieved, then so is the supremum of the other side.

PROOF OF LEMMA 3: Given any feasible u for the right-hand side of Lemma 3, we observe that $\phi = \psi = u$ is feasible for the left-hand side, and therefore the left-hand side is at least as large as the right-hand side. It therefore suffices to prove the reverse direction of the inequality. Let ϕ and ψ be feasible for the left-hand side. Given ϕ , it is clear that ψ must satisfy $\psi(y) \geq \sup_x [\phi(x) - \|x - y\|_1]$.

Set $\bar{\psi}(y) = \sup_x [\phi(x) - \|x - y\|_1]$. Since ψ exists, this supremum indeed has finite value. Since $\bar{\psi} \leq \psi$ pointwise, it follows that $\int_X \bar{\psi} d\mu_- \leq \int_X \psi d\mu_-$. We must now prove that $\bar{\psi} \in \mathcal{U}(X)$, thereby showing that $\phi, \bar{\psi}$ is feasible for the left-hand side and that replacing ψ by $\bar{\psi}$ does not decrease the objective value.

Claim 2. $\bar{\psi} \in \mathcal{U}(X)$ and $\bar{\psi} \in \mathcal{L}_1(X)$.

Proof. We will first show that $\bar{\psi} \in \mathcal{U}(X)$. We need to show continuity, monotonicity, and convexity.

- **Continuity.** Continuity of $\bar{\psi}$ follows from the Maximum Theorem since both ϕ and $\|\cdot\|_1$ are uniformly continuous.
- **Monotonicity.** Let $y \leq y'$ coordinate-wise and let x be arbitrary. We must show that there exists an x' such that $\phi(x) - \|x - y\|_1 \leq \phi(x') - \|x' - y'\|_1$. Set $x'_i = \max\{x_i, y'_i\}$. Since $x \leq x'$, we have $\phi(x) \leq \phi(x')$. We notice that if $x_i \geq y'_i$ then $x'_i = x_i$ and thus $|x'_i - y'_i| \leq |x_i - y_i|$, while if $x_i \leq y'_i$ then $|x'_i - y'_i| = 0$.

Therefore, we have that $\|x - y\|_1 \geq \|x' - y'\|_1$ and thus $\phi(x) - \|x - y\|_1 \leq \phi(x') - \|x' - y'\|_1$, as desired.

- **Convexity.** Let y, y', y'' be collinear points in X such that $y = \frac{y' + y''}{2}$. Then, given any x , we must show that there exist x' and x'' such that

$$\phi(x') - \|x' - y'\|_1 + \phi(x'') - \|x'' - y''\|_1 \geq 2\phi(x) - 2\|x - y\|_1.$$

We define x'_i and x''_i as follows:

- If $y'_i \geq y''_i$, set $x'_i = \max\{x_i, y'_i\}$ and $x''_i = \max\{2x_i - x'_i, y''_i\}$.
- If $y'_i < y''_i$, set $x''_i = \max\{x_i, y'_i\}$ and $x'_i = \max\{2x_i - x''_i, y''_i\}$.

Notice that $x' + x'' \geq 2x$, and thus (since ϕ is convex and monotone) we have $\phi(x') + \phi(x'') \geq 2\phi(x)$.

Suppose without loss of generality that $y'_i \geq y''_i$. We now consider two cases:

- $y'_i \geq x_i$. We then have $x'_i = y'_i$ and $x''_i = \max\{2x_i - y'_i, y''_i\}$. Therefore, $|y'_i - x'_i| = 0$ and $|y''_i - x''_i| \leq |y''_i - 2x_i + y'_i| = 2|y_i - x_i|$ since $y'_i + y''_i = 2y_i$.
- $y'_i < x_i$. We now have $x'_i = x_i$ and $x''_i = \max\{x_i, y''_i\} = x_i$. Therefore $|y''_i - x''_i| + |y'_i - x'_i|$ is equal to $|y''_i + y'_i - 2x_i|$, which equals $|2y_i - 2x_i|$.

Therefore, we have that $|y'_i - x'_i| + |y''_i - x''_i| \leq |2y_i - 2x_i|$ for all i , which implies that $\|x' - y'\|_1 + \|x'' - y''\|_1 \leq 2\|x - y\|_1$.

We have thus shown that $\bar{\psi} \in \mathcal{U}(X)$. We will now show that $\bar{\psi} \in \mathcal{L}_1(X)$. We have

$$\begin{aligned} \bar{\psi}(x) - \bar{\psi}(y) &= \sup_z \inf_w (\phi(z) - \|z - x\|_1 - \phi(w) + \|w - y\|_1) \\ &\leq \sup_z (\phi(z) - \|z - x\|_1 - \phi(z) + \|z - y\|_1) \\ &= \sup_z (\|z - y\|_1 - \|z - x\|_1) \leq \|x - y\|_1. \end{aligned}$$

□

Since $\phi, \bar{\psi}$ are a feasible pair of functions for the left-hand side of Lemma 3, we know that ϕ satisfies the inequality $\phi(x) \leq \inf_y [\bar{\psi}(y) + \|x - y\|_1]$. We now set $\bar{\phi}(x) = \inf_y [\bar{\psi}(y) + \|x - y\|_1]$. It is clear that the value of the left-hand objective function under $\bar{\phi}, \bar{\psi}$ is at least as large as its value under $\phi, \bar{\psi}$.

We claim that not only is $\bar{\phi}$ continuous, monotonic, and convex, but in fact that $\bar{\phi} = \bar{\psi}$. We notice that $\bar{\phi}(x) \leq \bar{\psi}(x) + \|x - x\|_1 = \bar{\psi}(x)$. To prove the other direction of the inequality, we compute

$$\bar{\phi}(x) = \inf_y [\bar{\psi}(y) + \|x - y\|_1] = \bar{\psi}(x) + \inf_y [\bar{\psi}(y) - \bar{\psi}(x) + \|x - y\|_1] \geq \bar{\psi}(x)$$

where the last inequality holds since $\bar{\psi}(x) - \bar{\psi}(y) \leq \|x - y\|_1$. Therefore $\bar{\phi} = \bar{\psi}$, and thus $\bar{\phi} \in \mathcal{U}(X)$. Since $\bar{\phi}$ satisfies the inequality $\bar{\phi}(x) - \bar{\phi}(y) \leq \|x - y\|_1$ it is feasible for the right-hand side of Lemma 3, and the value of the right-hand objective under $\bar{\phi}$ is at least as large the value of the left-hand objective under $\phi, \bar{\psi}$. We notice finally that if $\phi, \bar{\psi}$ are optimal for the left-hand side, then $\bar{\phi}$ is optimal for the right-hand side. \square

4.4.3 Proof of Theorem 2

By combining Lemma 1, Lemma 2, and Lemma 3, we have

$$\begin{aligned} & \inf_{\substack{\gamma \in \Gamma_+(X \times X) \\ \gamma_1 - \gamma_2 \preceq_1 \mu}} \int_{X \times X} \|x - y\|_1 d\gamma \geq \sup_{u \in \mathcal{U}(X) \cap \mathcal{L}_1(X)} \int_X u d\mu \\ & = \sup_{\substack{\phi, \psi \in \mathcal{U}(X) \\ \phi(x) - \psi(y) \leq \|x - y\|_1}} \left(\int_X \phi d\mu_+ - \int_X \psi d\mu_- \right) = \inf_{\substack{\gamma \in \Gamma_+(X \times X) \\ \gamma_1 \succeq_{cvx} \mu_+ \\ \gamma_2 \succeq_{cvx} \mu_-}} \int_{X \times X} \|x - y\|_1 d\gamma(x, y). \end{aligned}$$

By Lemma 2, the last minimization problem above achieves its infimum for some γ^* . We notice that γ^* is also feasible for the first minimization problem above, and therefore the inequality is actually an equality and γ^* is optimal for the first minimization problem. In addition, since γ^* is feasible for the last minimization problem, it satisfies $\gamma_1^*(X) = \gamma_2^*(X) = \mu_+(X)$.

To complete the proof of Theorem 2, we now show that the supremum of the maximization problem of Theorem 2 is achieved for some u^* . Consider a sequence of feasible functions $u_1, u_2, \dots \in \mathcal{U}(X) \cap \mathcal{L}_1(X)$ such that $\int_X u_i d\mu$ converges monotonically to the supremum value V , which we have proven is finite.⁶ Since $\mu(X) = 0$, we may without loss of generality assume that $u_i(0^n) = 0$ for all u_i . Since all of the functions are bounded by $\|x^{\text{high}}\|_1$ and are 1-Lipschitz (which implies equicontinuity), the Arzelà-Ascoli theorem implies that there exists a uniformly converging subsequence. Let u^* be the limit of that subsequence. Since the convergence is uniform, the function u^* is 1-Lipschitz, non-decreasing and convex and thus feasible for the mechanism design problem. Moreover, since the objective is linear, the revenue of the mechanism with that utility is equal to V and thus the supremum is achieved.

⁶Finiteness is also obvious because X is bounded and the infimum problem is feasible.

Chapter 5

Characterization of Optimal Mechanisms

5.1 Optimal Finite-Menu Mechanisms

To prove the optimality of our mechanisms in the examples of Section 4.3, we explicitly constructed a measure γ *separately* for each subset of types enjoying the same allocation in the optimal mechanism, establishing that the conditions of Corollary 1 are satisfied for each such subset of types separately. In this section, we show that decomposing the solution γ of the optimal transportation dual of Theorem 2 into “regions” of types enjoying the same allocation in the optimal solution u of the primal, and working on these regions separately to establish the complementary slackness conditions of Corollary 1 is guaranteed to work.

Even with this understanding of the structure of dual witnesses, it may still be non-trivial work to identify a witness certifying the optimality of a given mechanism. We thus develop a more usable framework for certifying the optimality of mechanisms, which does not involve finding dual witnesses at all. In particular, we show in Theorem 3 that a given mechanism \mathcal{M} is optimal for some f *if and only if* appropriate stochastic dominance conditions are satisfied by the restriction of the transformed measure μ of Definition 3 to each region of types enjoying the same allocation under \mathcal{M} . We thus provide conditions that are both *necessary* and *sufficient* for a given

mechanism \mathcal{M} to be optimal, a characterization result.

To describe our characterization, we define the intuitive notion a “menu” that a certain mechanism offers.

Definition 7. *The menu of a mechanism $\mathcal{M} = (\mathcal{P}, \mathcal{T})$ is the set*

$$\text{Menu}_{\mathcal{M}} = \{(p, t) : \exists x \in X, (p, t) = (\mathcal{P}(x), \mathcal{T}(x))\}.$$

Clearly, an IC mechanism allocates to every type x the option in the menu that maximizes that type’s utility. Figure 5-1 shows an example of a menu and the corresponding partition of the type set into subsets of types that prefer each option in the menu.

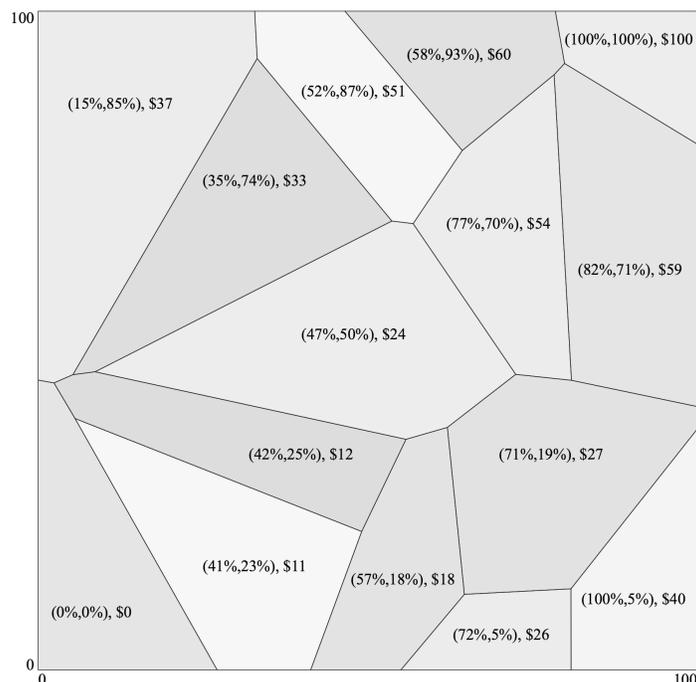


Figure 5-1: Partition of the type set $X = [0, 100]^2$ induced by some menu of lotteries.

The revenue of a mechanism with a finite menu-size comes from choices in the menu that are bought with strictly positive probability. The menu might contain options that are only bought with probability 0, but we can get another mechanism that gives identical revenue by removing all those options. We call this the *essential form* of a mechanism.

Definition 8. A mechanism \mathcal{M} is in essential form if for all options $(p, t) \in \text{Menu}_{\mathcal{M}}$, $\Pr_f[\{x \in X : (p, t) = (\mathcal{P}(x), \mathcal{T}(x))\}] > 0$.

We will now show our main result of this section under the assumption that the menu size is finite. We expect that our tools can be used to extend the results to the case of infinite menu size with a more careful analysis. We stress that the point of our result is not to provide sufficient conditions to certify optimality of mechanisms, as in [MV06, DDT13, GK14], but to provide necessary and sufficient conditions. In particular, we show that verifying optimality is *equivalent* to checking a collection of measure-theoretic inequalities, and this applies to arbitrary mechanisms with a finite menu-size. The most crucial component of the proof establishes that the optimal dual solution γ in Theorem 2 never convexly shuffles mass across regions of types that enjoy different allocations. (I.e. to obtain $\mu' = \gamma_1 - \gamma_2$ from μ we never need to move mass across different regions.) Similarly, we argue that the optimal γ never transports mass across regions.

Before formally stating our result, it is helpful to provide some intuition behind it. Consider a region R corresponding to a menu choice (\vec{p}, t) of an optimal mechanism \mathcal{M} . As we have already discussed, we can establish that the dual witness γ , which witnesses the optimality of \mathcal{M} , does not transport mass between regions and, likewise, the associated “convex shuffling” transforming μ to $\mu' = \gamma_1 - \gamma_2$ doesn’t shuffle across regions. Given this, our complementary slackness conditions of Corollary 1 imply then that $\mu_+|_R$ can be transformed to $\mu_-|_R$ using the following (intra-region R) operations:

- spreading positive mass within R so that the mean is preserved
- sending (positive) mass from a point $x \in R$ to a coordinatewise larger point $y \in R$ if for all coordinates where $y_i > x_i$ we have that the corresponding probability of the menu choice satisfies $p_i = 0$
- sending (positive) mass from a point $x \in R$ to a coordinatewise smaller point $y \in R$ if for all coordinates where $y_i < x_i$ we have that $p_i = 1$

Our characterization result involves stochastic dominance conditions that are

slightly more general than the standard notions of first, second and convex dominance. We need the following definition, which extends the notion of convex dominance.

Definition 9. We say that a function $u : X \rightarrow \mathbb{R}$ is \vec{v} -monotone for a vector $\vec{v} \in \{-1, 0, +1\}^n$ if it is non-decreasing in all coordinates i for which $v_i = 1$ and non-increasing in all coordinates i for which $v_i = -1$.

A measure α convexly dominates a measure β with respect to a vector $\vec{v} \in \{-1, 0, +1\}^n$, denoted $\alpha \succeq_{\text{cvx}(\vec{v})} \beta$, if for all convex \vec{v} -monotone functions $u \in \mathcal{U}(X)$:

$$\int u d\alpha \geq \int u d\beta.$$

Similarly, for vector random variables A and B with values in X , we say that $A \succeq_{\text{cvx}(\vec{v})} B$ if $\mathbb{E}[u(A)] \geq \mathbb{E}[u(B)]$ for all convex \vec{v} -monotone functions $u \in \mathcal{U}(X)$.

The definition of convex dominance presented earlier coincides with convex dominance with respect to the vector $\vec{1}$. Moreover, convex dominance with respect to the vector $-\vec{1}$ is related to second-order stochastic dominance as follows:

$$\alpha \succeq_{\text{cvx}(-\vec{1})} \beta \Leftrightarrow \beta \succeq_2 \alpha.$$

Measures satisfying the dominance condition of Definition 9 must have equal mass.

Proposition 1. Fix two measures $\alpha, \beta \in \Gamma(X)$ and a vector $v \in \{-1, 0, 1\}^n$. If it holds that $\alpha \succeq_{\text{cvx}(\vec{v})} \beta$, then $\alpha(X) = \beta(X)$.

We are now ready to describe our main characterization theorem. Our characterization, stated below as Theorem 3 and proven in the Section 5.3, is given in terms of the conditions of Definition 10.

Definition 10 (Optimal Menu Conditions). A mechanism \mathcal{M} satisfies the optimal menu conditions with respect to μ if for all menu choices $(p, t) \in \text{Menu}_{\mathcal{M}}$ we have

$$\mu_+|_R \succeq_{\text{cvx}(\vec{v})} \mu_-|_R$$

where $R = \{x \in X : (\mathcal{P}(x), \mathcal{T}(x)) = (p, t)\}$ is the subset of types that receive (p, t) and \vec{v} is the vector whose i -th coordinate v_i takes value 1 if $p_i = 0$, value -1 if $p_i = 1$ or value 0 if $p_i \in (0, 1)$.

Theorem 3 (Optimal Menu Theorem). *Let μ be the transformed measure of a probability density f as per Definition 3. Then a mechanism \mathcal{M} with finite menu size is an optimal IC and IR mechanism for a single additive buyer whose values for n goods are distributed according to the joint distribution f **if and only if** its essential form satisfies the optimal menu conditions with respect to μ .*

Interpretation of the Optimal Menu Conditions: A simple interpretation of the optimal menu conditions that Theorem 3 claims are necessary and sufficient for the optimality of mechanisms is this. Take some region R of the type set X corresponding to the types that are allocated a specific menu choice (p, t) by optimal mechanism \mathcal{M} . Let us consider the revenue $\int_R u^* d\mu$ extracted by \mathcal{M} from the types in region R . Is it possible to extract more revenue from these types? We claim that the optimal menu condition for region R guarantees that no mechanism can possibly extract more from the types in region R . Indeed, consider any utility function u induced by some other mechanism. The revenue extracted by this other mechanism in region R is $\int_R u d\mu = \int_R u^* d\mu + \int_R (u - u^*) d\mu \leq \int_R u^* d\mu$. That $\int_R (u - u^*) d\mu \leq 0$ follows directly from the optimal menu condition for region R . Indeed, since $u^*(x) = p \cdot x - t$ in region R , it follows that, whatever choice of u we made, $u - u^*$ is a convex \vec{v} -monotone function in region R , where \vec{v} is the vector defined by p as per Definition 10. Our condition in region R reads $\mu|_R \preceq_{\text{cvx}(\vec{v})} 0$, hence $\int_R (u - u^*) d\mu \leq 0$. Our line of argument implies the sufficiency of the optimal menu conditions, as they imply that for each region separately no mechanism can beat the revenue extracted by \mathcal{M} . The more surprising part (and harder to prove) is that the conditions are also necessary, implying that optimal mechanisms are locally optimal for every region R of types that they allocate the same menu choice to.

5.2 Grand Bundling Optimality and Applications

A particularly simple special case of our characterization result, pertains to the optimality of the grand-bundling mechanism. Theorem 3 implies that the mechanism that offers the grand bundle at price p is optimal *if and only if* the transformed measure μ satisfies a pair of stochastic dominance conditions. In particular, we obtain the following theorem:

Theorem 4 (Grand Bundling Optimality). *For a single additive buyer whose values for n goods are distributed according to the joint distribution f , the mechanism that only offers the bundle of all items at price p is optimal if and only if the transformed measure μ of f satisfies $\mu|_{\mathcal{W}} \succeq_2 0 \succeq_{\text{cvx}} \mu|_{\mathcal{Z}}$, where \mathcal{W} is the subset of types that can afford the grand bundle at price p , and \mathcal{Z} the subset of types who cannot.*

Next, we explore implications of our characterization of grand bundling optimality.

5.2.1 Application I: Selling items with large values

We now present an example application of our characterization result to determine the optimality of mechanisms that make a take-it-or-leave-it offer of the grand bundle of all items at some price. Our result applies to a setting with arbitrarily many items, which is relatively rare in the literature. More specifically, we consider a setting with n iid goods whose values are uniformly distributed on $[c, c + 1]$. It is easy to see that the ratio of the revenue achievable by grand bundling to the social welfare goes to 1 when either n or c goes to infinity.¹ This implies that grand-bundling is optimal or close to optimal for large values of n and c . Indeed, the following theorem shows that, for every n , grand bundling is the optimal mechanism for large values of c .

Theorem 5. *For any integer $n > 0$ there exists a c_0 such that for all $c \geq c_0$, the optimal mechanism for selling n iid goods whose values are uniform on $[c, c + 1]$ is a take-it-or-leave-it offer for the grand bundle.*

¹This follows by setting a price for the grand-bundle equal to $(c + \frac{1}{2})n - \sqrt{n} \log cn$ and noting that a straightforward application of Hoeffding's inequality gives that the bundle is accepted with probability close to 1.

Remark 3. [Pav11] proved the above result for two items, and explicitly solved for $c_0 \approx 0.077$. In our proof, for simplicity of analysis, we do not attempt to exactly compute c_0 as a function of n .

Our proof of Theorem 5 uses the following lemma, which enables us to appropriately match regions on the surface of a hypercube.

Lemma 4. For $n \geq 2$ and $\rho > 1$, define the $(n - 1)$ -dimensional subsets of $[0, 1]^n$:

$$A = \left\{ x : 1 = x_1 \geq x_2 \geq \cdots \geq x_n \text{ and } x_n \leq 1 - \left(\frac{\rho - 1}{\rho} \right)^{1/(n-1)} \right\}$$

$$B = \{ y : y_1 \geq \cdots \geq y_n = 0 \}.$$

There exists a continuous bijective map $\varphi : A \rightarrow B$ such that

- For all $x \in A$, x is componentwise greater than or equal to $\varphi(x)$
- For subsets $S \subseteq A$ which are measurable under the $(n - 1)$ -dimensional surface Lebesgue measure $v(\cdot)$, it holds that $\rho \cdot v(S) = v(\varphi(S))$.
- For all $\epsilon > 0$, if $\varphi_1(x) \leq \epsilon$ then $x_n \geq 1 - \left(\frac{\epsilon^{n-1} + \rho - 1}{\rho} \right)^{1/(n-1)}$.

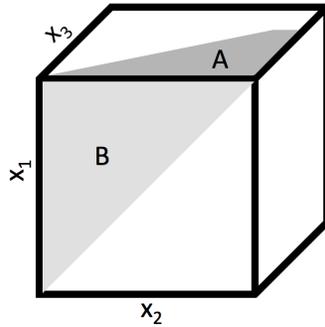


Figure 5-2: The regions of Lemma 4 for the case $n = 3$.

Proof. We define the mapping $\varphi : A \rightarrow B$ by $\varphi(x) = y$, where

$$y_1 = \left[1 - \rho \left(1 - (1 - x_n)^{n-1} \right) \right]^{1/(n-1)}; \quad y_i = \frac{x_i - x_n}{1 - x_n} \cdot y_1 \quad \text{for } i > 1.$$

We first claim that φ is a bijection. As x_n ranges from 0 to $1 - \left(\frac{\rho-1}{\rho}\right)^{1/(n-1)}$, we see that y_1 ranges from 1 to 0, and thus there is a bijection between valid y_1 values and valid x_n values. Furthermore, for any fixed y_1 and x_n , there is a bijection between x_i and y_i for $i = 2, \dots, n-1$. (By varying x_i between x_n and 1 we can achieve all values of y_i between 0 and y_1 .) Furthermore, for any fixed y_1 and x_n the mapping from x_i to y_i is an increasing function of x_i , and therefore for all $x \in A$ we have $y_1 \in [0, 1]$ and $y_1 \geq y_2 \geq \dots \geq y_n = 0$. Thus, φ is a bijection between A and B . Next, we claim that for any $x \in A$, it holds that x is componentwise at least as large as $\varphi(x)$. Since $x_1 = 1$, it trivially holds that $x_1 \geq \varphi_1(x)$. Fix a value of x_n (and hence of y_1), and consider the bijection $g : [x_n, 1] \rightarrow [0, y_1]$ given by $g(z) = y_1(z - x_n)/(1 - x_n)$. We must show that $z - g(z) \geq 0$ for all $z \in [x_n, 1]$. This follows from noticing that $z - g(z)$ is a linear function of z and both $x_n - g(x_n) = x_n$ and $1 - g(1) = 1 - y_1$ are nonnegative.

We now show that φ scales surface measure of every measurable $S \subset A$ by a factor of $1/\rho$. Instead of directly analyzing surface measures, it suffices to prove that the function $\varphi' : W \rightarrow W$ scales volumes by ρ , where $W \subset \mathbb{R}^{n-1}$ is the set $\{w : 1 \geq w_1 \geq \dots \geq w_{n-1} \geq 0\}$ and $\varphi'(w)$ drops the last (constant) coordinate of $\varphi(1, w_1, \dots, w_{n-1})$ and then (for notational convenience) permutes the first coordinate to the end. That is,

$$\varphi'(w_1, \dots, w_{n-1}) = \left(\frac{w_1 - w_{n-1}}{1 - w_{n-1}} z(w_{n-1}), \dots, \frac{w_{n-2} - w_{n-1}}{1 - w_{n-1}} z(w_{n-1}), z(w_{n-1}) \right)$$

where $z(w_{n-1}) = [1 - \rho(1 - (1 - w_{n-1})^{n-1})]^{1/(n-1)}$.

We now analyze the determinant of the Jacobian matrix J of φ' . We notice that the only non-zero entries of J are the diagonals and the rightmost column. In particular, J is upper triangular, and therefore its determinant is the product of its

diagonal entries. We therefore compute

$$\begin{aligned} \det(J) &= \left(\frac{z(w_{n-1})}{1-w_{n-1}} \right)^{n-2} \cdot \frac{\partial}{\partial w_{n-1}} [1 - \rho (1 - (1 - w_{n-1})^{n-1})]^{1/(n-1)} \\ &= \left(\frac{z(w_{n-1})}{1-w_{n-1}} \right)^{n-2} \cdot \frac{-1}{n-1} (z(w_{n-1})^{-(n-2)} \cdot \rho \cdot (n-1)(1-w_{n-1})^{n-2}) = -\rho \end{aligned}$$

as desired.

Lastly, suppose $y_1 \leq \epsilon$. Then $[1 - \rho(1 - (1 - x_n)^{n-1})]^{1/(n-1)} \leq \epsilon$ and thus $x_n \geq 1 - \left(\frac{\epsilon^{n-1} + \rho - 1}{\rho} \right)^{1/(n-1)}$. \square

With Lemma 4 in hand, we are now ready to prove Theorem 5. The main difficulty of the proof is verifying the necessary stochastic dominance relations above the grand bundling hyperplane. Our proof appropriately partitions this part of the hypercube into $2(n! + 1)$ regions and uses Lemma 4 to show a desired stochastic dominance relation holds for an appropriate pairing of regions.

PROOF OF THEOREM 5: Fix the number of items n . For any value of c , the transformed measure on the hypercube $(c, c + 1)^n$ we obtain is as follows:

- A point mass of +1 at (c, c, \dots, c) .
- Mass of $-(n + 1)$ uniformly distributed throughout the interior.
- Mass of $-c$ distributed on each surface $x_i = c$ of the hypercube.
- Mass of $c + 1$ distributed on each surface $x_i = c + 1$ of the hypercube.

For notational convenience when checking the stochastic dominance properties of Theorem 3, we will shift the hypercube to the origin. That is, we will consider instead the measure μ^c on $[0, 1]^n$ which has mass +1 at the origin, mass of $-c$ on each surface $x_i = 0$, et cetera. It is important to notice that the mass that μ assigns to the interior of $[0, 1]^n$ and to the origin do not depend on c , while the mass on each surface is a function of c .

For any $h \in (0, 1)$, define the region $Z(h) = \{x \in [0, 1]^n : \|x\|_1 \leq h\}$. For any fixed c_0 , it holds that $\mu_+^{c_0}(Z(h)) = 1$ for all $h \in (0, 1)$ and there exists a small enough $h' > 0$

such that $\mu_-^{c_0}(Z(h')) < 1$. Since for this fixed h' it holds that $\mu_-^c(Z(h'))$ increases with c (and becomes arbitrarily large as c becomes large), there must exist a $c' > c_0$ such that $\mu_-^{c'}(Z(h')) = 1$, and thus $\mu^{c'}(Z(h')) = 0$. We can therefore pick a decreasing function $p^* : \mathbb{R}_{\geq 0} \rightarrow (0, 1)$ such that, for all sufficiently large c , $\mu^c(Z(p^*(c))) = 0$.² As argued above, for any small enough $h' > 0$ there exists a c' such that $\mu_-^{c'}(Z(h')) = 1$ and thus $p^*(c') = h'$. It follows that $p^*(c) \rightarrow 0$ as $c \rightarrow \infty$.

For all c , define the following subsets of $[0, 1]^n$:

$$Z_c = \{x : \|x\|_1 \leq p^*(c)\}; \quad W_c = \{x : \|x\|_1 \geq p^*(c)\}.$$

We notice that $\mu_+^c(Z_c \cap W_c) = \mu_-^c(Z_c \cap W_c) = 0$. By construction, for large enough c we have $\mu^c(Z_c) = 0$. In addition, the only positive mass in Z_c is at the origin, and thus $\mu_-^c|_{Z_c} \succeq_{cvx} \mu_+^c|_{Z_c}$.

To apply Theorem 3, it remains to show that, for sufficiently large c , $\mu_+^c|_{W_c} \preceq_{cvx(-\bar{1})} \mu_-^c|_{W_c}$. To prove this, we partition W_c into $2(n! + 1)$ disjoint³ regions, $P_0, P_{\sigma_1}, \dots, P_{\sigma_{n!}}$ and $N_0, N_{\sigma_1}, \dots, N_{\sigma_{n!}}$, where σ_j is a permutation of $1, \dots, n$. This partition will be such that $\cup_j P_j$ contains the entire support of $\mu_+^c|_{W_c}$ and $\cup_j N_j$ contains the entire support of $\mu_-^c|_{W_c}$. We will show that $\mu_+^c|_{P_j} \preceq_{cvx(-\bar{1})} \mu_-^c|_{N_j}$ for all j , thereby proving $\mu_+^c|_{W_c} \preceq_{cvx(-\bar{1})} \mu_-^c|_{W_c}$.

For every permutation σ of $1, \dots, n$, define:

$$P'_\sigma = \left\{ x : 1 = x_{\sigma(1)} \geq x_{\sigma(2)} \geq \dots \geq x_{\sigma(n)} \geq 0 \text{ and } x_{\sigma(n)} \leq 1 - \left(\frac{1}{c+1} \right)^{1/(n-1)} \right\}$$

$$N'_\sigma = \{ y : 1 \geq y_{\sigma(1)} \geq \dots \geq y_{\sigma(n-1)} \geq y_{\sigma(n)} = 0 \}$$

Denote by $\rho \triangleq (c+1)/c$ the ratio between the surface densities of μ_+^c and μ_-^c on P'_σ and N'_σ , respectively, and let $\varphi_\sigma : P'_\sigma \rightarrow N'_\sigma$ be the bijection given by Lemma 4. By construction, $\mu_+^c(S) = \mu_-^c(\varphi_\sigma(S))$ for all measurable $S \subseteq P'_\sigma$.

²Our intention is to argue that for c large enough, the optimal mechanism will be grand bundling for a price of $p^*(c) + c$, where the additive $+c$ term comes from our shift of the hypercube to the origin.

³For notational simplicity, our regions overlap slightly, although the overlap always has zero mass under both μ_+^c and μ_-^c .

Denote $N_\sigma \triangleq N'_\sigma \setminus Z_c$ and $P_\sigma \triangleq \varphi^{-1}(N_\sigma)$. By construction, φ is a bijection between P_σ and N_σ , preserving the respective the measures μ_+^c and μ_-^c , such that for all $x \in P_\sigma$, x is componentwise at least as large as $\varphi(x)$. Therefore, by Strassen's theorem, $\mu_+^c|_{P_\sigma} \preceq_{cvx(-\bar{1})} \mu_-^c|_{N_\sigma}$. Lastly, we define

$$P_0 = \{x \in [0, 1]^n : x_i = 1 \text{ for some } i\} \setminus \left(\bigcup_\sigma P_\sigma \right); \quad N_0 = (0, 1)^n \setminus Z_c.$$

P_0 consists of all points on the outer surface of the hypercube which have not yet been matched to any N_σ , and N_0 consists of all points on which μ_-^c is nontrivial which have not yet been matched.⁴ It therefore remains only to show that $\mu_+^c|_{P_0} \preceq_{cvx(-\bar{1})} \mu_-^c|_{N_0}$.

We claim that, for large enough c , P_0 only contains points with all coordinates greater than $3/4$. Indeed:

- Every x with $x_i = 1$ but some $x_j < 1 - \left(\frac{1}{c+1}\right)^{1/(n-1)}$ is in some P'_σ .
- For large c , every x with $x_i = 1$ but some $x_j \leq 3/4$ is in some P'_σ .
- We claim that for large c , every $x \in P'_\sigma \setminus P_\sigma$ has all coordinates at least $3/4$. Indeed, for every $x \in P'_\sigma \setminus P_\sigma$, it must be that $\varphi(x) \in Z_c$, and thus $\|\varphi(x)\|_1 \leq p^*(c)$. By Lemma 4, we have $x_{\sigma(n)} \geq 1 - \left(\frac{p^*(c)^{n-1} + \rho - 1}{\rho}\right)^{1/(n-1)}$. As c gets large, $\rho \rightarrow 1$ and $p^*(c) \rightarrow 0$. Thus, for sufficiently large c , we have $x \in P'_\sigma \setminus P_\sigma$ implies $x_{\sigma(n)} \geq 3/4$. Since $x_{\sigma(n)}$ is the smallest coordinate of x , it follows that all coordinates of any $x \in P'_\sigma \setminus P_\sigma$ are greater than $3/4$.
- Thus, for sufficiently large c , every x with $x_i = 1$ but some $x_j < 3/4$ lies in some P_σ , and hence does not lie in P_0 .

By construction, $\mu_-^c|_{N_0}$ and $\mu_+^c|_{P_0}$ have the same total mass. Consider independent random variables X and Y corresponding to $\mu_-^c|_{N_0}$ and $\mu_+^c|_{P_0}$, respectively, where we scale both measures so that they are probability distributions. By Lemma 5, it suffices to show that for sufficiently large c , $Y \geq \mathbb{E}[X]$ almost surely.⁵ Since $\mu_+^c|_{P_0}$ is

⁴All other points on which μ_-^c is nontrivial have been matched either to the origin (if the point lies in Z_c), or to some point in P_σ (if the point lies in $N'_\sigma \setminus Z_c$).

⁵In general, to prove second order dominance we might need to nontrivially couple X and Y . In this case, however, choosing independent random variables suffices.

supported on P_0 , we need only show that all coordinates of $\mathbb{E}[X]$ are less than $3/4$. We recall that μ_-^c assigns a total mass of $n + 1$, distributed uniformly, to the interior of the hypercube. As c gets large, $p^*(c)$ approaches 0, and thus

$$\frac{\mu_-^c(Z_c \cap (0, 1)^n)}{\mu_-^c((0, 1)^n)} \rightarrow 0$$

For large c , therefore, $\mathbb{E}[X]$ becomes arbitrarily close to the center of the hypercube, which is the point with all coordinates equal to $1/2$. Therefore we have

$$\mu_+^c|_{P_0} \preceq_{cvx(-\bar{1})} \mu_-^c|_{N_0}$$

□

5.2.2 Application II: Selling a large number of items

We now revisit the previous example with n items with values distributed uniformly in $[c, c + 1]$ and consider what happens when n becomes large while c remains fixed. In this case, in contrast to the previous result, we show using our strong duality theorem that grand bundling is *never* the optimal mechanism for sufficiently large values of n .

Theorem 6. *For any $c \geq 0$ there exists an integer n_0 such that for all $n \geq n_0$, the optimal mechanism for selling n iid goods whose values are uniform on $[c, c + 1]$ is **not** a take-it-or-leave-it offer for the grand bundle.*

Proof. Given c , let n be large enough so that

$$\frac{n + 1}{n!} + \frac{nc}{(n - 1)!} < 1.$$

To prove the theorem, we will assume that an optimal grand bundling price p exists and reach a contradiction.

As shown in Section 3.1.3, under the transformed measure μ the hypercube has mass $-(n + 1)$ in the interior, $+1$ on the origin, $c + 1$ on every positive surface $x_i = c + 1$, and $-c$ on every negative surface $x_i = c$.

According to Theorem 3, for grand bundling at price p to be optimal it must hold that $\mu|_{Z_p} \preceq_{cvx} 0$ for the region $Z_p = \{x : \|x\|_1 \leq p\}$. If $p > nc + 1$ this could not happen, since for the function $\mathbb{1}_{x_1=c+1}(x)$ (which is increasing and convex in $[c, c+1]^n$) we have that $\int_{Z_p} \mathbb{1}_{x_1=c+1} d\mu = \mu(Z_p \cap \{x_1 = c + 1\}) = \mu_+(Z_p \cap \{x_1 = c + 1\}) > 0$ which violates the $\mu|_{Z_p} \preceq_{cvx} 0$ condition.

To complete the proof, we now consider the case that $p \leq nc + 1$ and will derive a contradiction. For the necessary condition $\mu|_{Z_p} \preceq_{cvx} 0$ to hold, it must be that $\mu(Z_p) = 0$. Since $p \leq nc + 1$, none of the positive outer surfaces of the cube have nontrivial intersection with Z_p , so all the positive mass in Z_p is located at the origin. Therefore, $\mu_+(Z_p) = 1$ which means that $\mu_-(Z_p) = 1$ as well. Moreover, since $p \leq nc + 1 \Rightarrow Z_p \subseteq Z_{nc+1}$, we also have that $\mu_-(Z_{nc+1}) \geq \mu_-(Z_p) = 1$.

To reach a contradiction, we will show that $\mu_-(Z_{nc+1}) < 1$. We observe that we can compute $\mu_-(Z_{nc+1})$ directly by summing the n -dimensional volume of the negative interior with the $(n - 1)$ -dimensional volumes of each of the n negative surfaces enclosed in Z_{nc+1} .⁶ The first is equal to:

$$\begin{aligned} (n + 1) \times \text{Vol}[\{x \in (c, c + 1)^n : \|x\|_1 \leq nc + 1\}] = \\ (n + 1) \times \text{Vol}[\{x \in (0, 1)^n : \|x\|_1 \leq 1\}] = \frac{(n + 1)}{n!} \end{aligned}$$

while the latter is equal to:

$$\begin{aligned} n \times c \times \text{Vol}[\{x \in (c, c + 1)^{n-1} : \|x\|_1 + c \leq nc + 1\}] = \\ n \times c \times \text{Vol}[\{x \in (0, 1)^{n-1} : \|x\|_1 \leq 1\}] = \frac{nc}{(n - 1)!} \end{aligned}$$

Therefore, we get that $1 \leq \mu_-(Z_{nc+1}) = \frac{(n+1)}{n!} + \frac{nc}{(n-1)!}$ which is a contradiction since we chose n to be sufficiently large to make this quantity less than 1. \square

⁶The geometric intuition of this step of the argument is that, for large enough n , the fraction of the n -dimensional hypercube $[0, 1]^n$ which lies below the diagonal $\|x\|_1 = 1$ goes to zero, and similarly the fraction of $(n - 1)$ -dimensional surface area on the boundaries which lies below the diagonal also goes to zero as n gets large.

5.3 Proof of Stochastic Conditions of Section 5.1

We complete this chapter by giving the proof of Theorem 3. We begin by presenting some useful probabilistic tools that will be essential for the proof.

5.3.1 Probabilistic Lemmas

We first present a useful result about convex dominance of random variables. For more information about this result, see Theorem 7.A.2 of [SS10].

Lemma 5 (Strassen's Theorem). *Let A and B be random vectors. Then $A \preceq_{cvx} B$ if and only if there exist random vectors \hat{A} and \hat{B} , defined on the same probability space, such that $\hat{A} =_{st} A$, $\hat{B} =_{st} B$, and $\mathbb{E}[\hat{B}|\hat{A}] \geq \hat{A}$ almost surely, where the final inequality is componentwise and where $=_{st}$ denotes equality in distribution.*

It is easy to extend the above result to convex dominance with respect to a vector \vec{v} as defined in Definition 9.

Lemma 6 (Extended Strassen's Theorem). *Let A and B be random vectors. Then $A \preceq_{cvx(\vec{v})} B$ if and only if there exist random vectors \hat{A} and \hat{B} , defined on the same probability space, with $\hat{A} =_{st} A$, $\hat{B} =_{st} B$, such that (almost surely):*

- if $v_i = +1$, then $E[\hat{B}_i|\hat{A}] \geq \hat{A}_i$
- if $v_i = 0$, then $E[\hat{B}_i|\hat{A}] = \hat{A}_i$
- if $v_i = -1$, then $E[\hat{B}_i|\hat{A}] \leq \hat{A}_i$

We now state a multivariate variant of Jensen's inequality along with the necessary condition for equality to hold. The proof of this result is standard and straightforward, and thus is omitted.

Lemma 7 (Jensen's inequality). *Let V be a vector-valued random variable with values in $[0, M]^n$ and let u be a convex Lipschitz-continuous function mapping $[0, M]^n \rightarrow \mathbb{R}$. Then $\mathbb{E}[u(V)] \geq u(\mathbb{E}[V])$. Furthermore, equality holds if and only if, for every a in the subdifferential of u at $\mathbb{E}[V]$, the equality $u(V) = a \cdot (V - \mathbb{E}[V]) + u(\mathbb{E}[V])$ holds almost surely.*

The following lemma is a conditional variant of Lemma 7, based on the multivariate conditional Jensen's inequality, as in Theorem 10.2.7 of [Dud02]. This lemma is used as a tool for Lemma 9, the main result of this subsection.

Lemma 8. *Let (Ω, \mathcal{A}, P) be a probability space, V be a random variable on Ω with values in X where $X = \prod_{i=1}^n [x_i^{low}, x_i^{high}]$, and $u : X \rightarrow \mathbb{R}$ be convex and Lipschitz continuous. Let \mathcal{C} be any sub- σ -algebra of \mathcal{A} and suppose that $\mathbb{E}[u(V)|\mathcal{C}] = u(\mathbb{E}[V|\mathcal{C}])$ almost-surely. Then for almost all $x \in \Omega$ the equality $u(y) = a_{y_x} \cdot (y - y_x) + u(y_x)$ holds almost surely with respect to the law ⁷ $P_{V|\mathcal{C}}(\cdot, x)$, where y_x is the expectation of the random variable with law $P_{V|\mathcal{C}}(\cdot, x)$ and a_{y_x} is any subgradient of u at y_x .*

PROOF OF LEMMA 8: The proof is based on the proof of the multivariate conditional Jensen's inequality, as in Theorem 10.2.7 of [Dud02]. This theorem requires $|V|$ and $u \circ V$ to be integrable, which is true in our setting. We note that the theorem applies when u is defined in an open convex set, but because u is Lipschitz continuous we can extend it to a function with domain an open set containing X . The multivariate conditional Jensen's inequality states that, almost surely, $\mathbb{E}[V|\mathcal{C}] \in \mathcal{C}$ and $\mathbb{E}[u(V)|\mathcal{C}] \geq u(\mathbb{E}[V|\mathcal{C}])$. The proof of Theorem 10.2.7 in [Dud02] furthermore shows that the following two equalities hold:

$$\mathbb{E}[V|\mathcal{C}](x) = \int_X y P_{V|\mathcal{C}}(dy, x); \quad \mathbb{E}[u(V)|\mathcal{C}](x) = \int_X u(y) P_{V|\mathcal{C}}(dy, x).$$

Since $\mathbb{E}[u(V)|\mathcal{C}](x) = u(\mathbb{E}[V|\mathcal{C}](x))$ for almost all x , we apply the unconditional Jensen inequality (Lemma 7) to the laws $P_{V|\mathcal{C}}(\cdot, x)$ to prove the lemma. \square

We now present Lemma 9. This lemma states that for random variables A and B with $A \preceq_{cvx} B$ if it holds that $u(A) = u(B)$ for some convex function u , then there exists a coupling between A and B with several desirable properties, including that points are only matched if u shares a subgradient at these points.

Lemma 9. *Let A and B be vector random variables with values in X , where $X = \prod_{i=1}^n [x_i^{low}, x_i^{high}]$, such that $A \preceq_{cvx} B$. Let $u : X \rightarrow \mathbb{R}$ be 1-Lipschitz with respect to the*

⁷The law $P_{V|\mathcal{C}}(\cdot, x)$ allows us to express the conditional distribution of V given \mathcal{C}

ℓ_1 norm, convex, and monotonically non-decreasing. Suppose that $\mathbb{E}[u(A)] = \mathbb{E}[u(B)]$ and that $g : X \rightarrow [0, 1]^n$ is a measurable function such that for all $z \in X$, $g(z)$ is a subgradient of u at z .

Then there exist random variables $\hat{A} =_{st} A$ and $\hat{B} =_{st} B$ such that, almost surely:

- $u(\hat{B}) = u(\hat{A}) + g(\hat{A}) \cdot (\hat{B} - \hat{A})$
- $g(\hat{A})$ is a subgradient of u at \hat{B} .
- $\mathbb{E}[\hat{B}|\hat{A}]$ is componentwise greater or equal to \hat{A}
- $u(\mathbb{E}[\hat{B}|\hat{A}]) = u(\hat{A})$.

PROOF OF LEMMA 9: By Lemma 5, there exist random variables $\hat{A} =_{st} A$ and $\hat{B} =_{st} B$ such that $\mathbb{E}[\hat{B}|\hat{A}]$ is componentwise greater than or equal to \hat{A} almost surely. We have

$$0 = \mathbb{E}[u(\hat{B}) - u(\hat{A})] \geq \mathbb{E}[u(\hat{B}) - u(\mathbb{E}[\hat{B}|\hat{A}])] = \mathbb{E}[\mathbb{E}[u(\hat{B})|\hat{A}] - u(\mathbb{E}[\hat{B}|\hat{A}])] \geq 0$$

and therefore $\mathbb{E}[\mathbb{E}[u(\hat{B})|\hat{A}]] = \mathbb{E}[u(\mathbb{E}[\hat{B}|\hat{A}])] = \mathbb{E}[u(\hat{B})] = \mathbb{E}[u(\hat{A})]$.

Since u is monotonic, $u(\hat{A}) \leq u(\mathbb{E}[\hat{B}|\hat{A}])$ almost surely. Since $\mathbb{E}[u(\hat{A})] = \mathbb{E}[u(\mathbb{E}[\hat{B}|\hat{A}])]$, it follows that $u(\hat{A}) = u(\mathbb{E}[\hat{B}|\hat{A}])$ almost surely.

Select any collection of random variables $\{\hat{B}_{|\hat{A}=x}\}$ corresponding to the laws $P_{\hat{B}|\hat{A}}(\cdot, x)$. For almost all values x of \hat{A} , $\mathbb{E}[\hat{B}_{|\hat{A}=x}]$ is componentwise greater than x and $u(x) = u(\mathbb{E}[\hat{B}_{|\hat{A}=x}])$. We claim now that any subgradient a_x of u at x is also a subgradient of u at $\mathbb{E}[\hat{B}_{|\hat{A}=x}]$. Indeed, choose such a subgradient a_x . We compute

$$u(\mathbb{E}[\hat{B}_{|\hat{A}=x}]) \geq u(x) + a_x \cdot (\mathbb{E}[\hat{B}_{|\hat{A}=x}] - x) = u(\mathbb{E}[\hat{B}_{|\hat{A}=x}]) + a_x \cdot (\mathbb{E}[\hat{B}_{|\hat{A}=x}] - x)$$

and therefore $a_x \cdot \mathbb{E}[\hat{B}_{|\hat{A}=x}] = a_x \cdot x$, by non-negativity of the subgradient. Furthermore, for any point $z \in X$,

$$\begin{aligned} u(z) &\geq u(x) + a_x \cdot (z - x) = u(\mathbb{E}[\hat{B}_{|\hat{A}=x}]) + a_x \cdot (z - x) \\ &= u(\mathbb{E}[\hat{B}_{|\hat{A}=x}]) + a_x \cdot (z - \mathbb{E}[\hat{B}_{|\hat{A}=x}]) \end{aligned}$$

and thus a_x is a subgradient of u at $\mathbb{E}[\hat{B}|_{\hat{A}=x}]$.

Since $\mathbb{E}[\mathbb{E}[u(\hat{B})|\hat{A}]] = \mathbb{E}[u(\mathbb{E}[\hat{B}|\hat{A}])]$, by Jensen's inequality it follows that $\mathbb{E}[u(\hat{B})|\hat{A}] = u(\mathbb{E}[\hat{B}|\hat{A}])$ almost surely. By Lemma 8, it therefore holds for almost all values x of \hat{A} that the equality

$$\begin{aligned} u(y) &= a_x \cdot (y - \mathbb{E}[\hat{B}|_{\hat{A}=x}]) + u(\mathbb{E}[\hat{B}|_{\hat{A}=x}]) = a_x \cdot (y - x) + u(\mathbb{E}[\hat{B}|_{\hat{A}=x}]) \\ &= a_x \cdot (y - x) + u(x) \end{aligned}$$

holds $\hat{B}|_{\hat{A}=x}$ almost surely.

Lastly, we will show that, almost surely, a_x is a subgradient of u at $\hat{B}|_{\hat{A}=x}$. Indeed, for any $p \in X$, and almost all values of x we have

$$\begin{aligned} u(p) &\geq u(x) + a_x \cdot (p - x) = u(x) + a_x \cdot (\hat{B}|_{\hat{A}=x} - x) + a_x \cdot (p - \hat{B}|_{\hat{A}=x}) \\ &= u(\hat{B}|_{\hat{A}=x}) + a_x \cdot (p - \hat{B}|_{\hat{A}=x}). \end{aligned}$$

□

5.3.2 Proof of the Optimal Menu Theorem (Theorem 3)

To prove the equivalence we prove both implications of the theorem separately.

Sufficiency Conditions

We will show that the Optimal Menu Conditions of Definition 10 imply that a mechanism \mathcal{M} is optimal. To show the theorem, we construct a measure γ such that the conditions of Corollary 1 are satisfied. We will construct this measure separately for every region that corresponds to a menu choice of mechanism \mathcal{M} .

Consider a menu choice $(p, t) \in \text{Menu}_{\mathcal{M}}$, the corresponding region R and the corresponding vector \vec{v} as in Definition 10. Let A and B be random vectors distributed according to the (normalized) measures $\mu_+|_R$ and $\mu_-|_R$. From the Optimal Menu Conditions, we have that $A|_R \preceq_{cvx(\vec{v})} B|_R$ (almost surely). By the extended version of Strassen's theorem (Lemma 6), it holds that there exist random vectors \hat{A}, \hat{B} with

$\hat{A} =_{st} A|_R$ and $\hat{B} =_{st} B|_R$, such that (almost surely):

- if $v_i = +1$, then $E[\hat{B}_i|\hat{A}] \geq \hat{A}_i$
- if $v_i = 0$, then $E[\hat{B}_i|\hat{A}] = \hat{A}_i$
- if $v_i = -1$, then $E[\hat{B}_i|\hat{A}] \leq \hat{A}_i$

Now define the random variable $\hat{C} = \min(E[\hat{B}|\hat{A}], \hat{A})$ where we take the coordinate-wise minimum. We now have that (almost surely):

- if $v_i = +1$, then $E[\hat{B}_i|\hat{A}] \geq \hat{A}_i = \hat{C}_i$
- if $v_i = 0$, then $E[\hat{B}_i|\hat{A}] = \hat{A}_i = \hat{C}_i$
- if $v_i = -1$, then $\hat{C}_i = E[\hat{B}_i|\hat{A}] \leq \hat{A}_i$

Let γ_R be the measure according to which the vector (\hat{A}, \hat{C}) is distributed. By construction, $\gamma_{R1} = \mu_+|_R$ and $\gamma_{R2} \preceq_{cvx} \mu_-|_R$, and thus $\gamma_{R1} - \gamma_{R2} \succeq_{cvx} \mu|_R$. Moreover, the conditions of Corollary 1 are satisfied:

- $u(x) - u(y) = \|x - y\|_1$, is satisfied $\gamma_R(x, y)$ -almost surely since \hat{A} is larger than \hat{C} only in coordinates for which $v_i = -1$ and thus $p_i = 1$.
- $\int ud(\gamma_{R1} - \gamma_{R2}) = \int ud(\mu_+|_R - \mu_-|_R)$ is satisfied: By definition we have that $\int ud\gamma_{R1} = \int ud\mu_+|_R$. Moreover, we can also show that $\int ud\gamma_{R2} = \int ud\mu_-|_R$ by noting that $\int ud\mu_-|_R = \mu_-(R)E[u(\hat{B})] = \mu_-(R)E[p \cdot \hat{B} - t] = \mu_-(R)E[p \cdot E[\hat{B}|\hat{A}] - t]$ and that $\mu_-(R)E[p \cdot E[\hat{B}|\hat{A}] - t]$ is equal to $\mu_-(R)E[p \cdot \hat{C} - t] = \int ud\gamma_{R2}$ since $\hat{C}_i \neq E[\hat{B}_i|\hat{A}]$ only when $E[\hat{B}_i|\hat{A}]$ is strictly larger than \hat{A}_i which only happens only in coordinates i where $v_i = +1$ and thus $p_i = 0$.

This completes the proof that the Optimal Menu Conditions imply optimality of the mechanism since we can construct a feasible measure γ satisfying the conditions of Corollary 1 by considering the sum of the constructed measures for each region.

Optimality implies Stochastic Conditions

We will now prove the other direction of the result. Consider an optimal mechanism $\mathcal{M} = (\mathcal{P}, \mathcal{T})$ with a finite menu size over type space $X = \prod_{i=1}^n [x_i^{\text{low}}, x_i^{\text{high}}]$. Since \mathcal{M} is given in essential form, in the menu of \mathcal{M} there is no dominated option. So for all options on the menu there is a set of buyer types that strictly prefer it from any other option, and that set of types occurs with positive probability.

Now, define the set $Z = \{x \in X : p \cdot x - t = \mathcal{P}(x) \cdot x - \mathcal{T}(x) \text{ for } (p, t) \in \text{Menu}_{\mathcal{M}} \text{ with } (p, t) \neq (\mathcal{P}(x), \mathcal{T}(x))\}$. This is the set of types where there is no single option that is the best and it is where the utility function of the mechanism is not differentiable. We show the following lemma.

Lemma 10. $\mu_-(Z) = 0$

Proof. Note that, by its construction, μ_- assigns zero mass to any k -dimensional surface for $k \leq n - 2$. Moreover, it only assigns mass to $(n - 1)$ -dimensional surfaces which lie along the boundary of X .

Every pair of distinct choices $(p, t), (p', t') \in \text{Menu}_{\mathcal{M}}$ defines a hyperplane $p \cdot x - t = p' \cdot x - t'$ containing the types who derive the same utility from these two choices. As the menu is finite, there exist a finite number of such pairs, hence a finite number of hyperplanes. The set Z contains a subset of types in the finite union of these hyperplanes, so μ_- assigns no mass to the subset of Z which lies on the interior of X .

Regarding the μ_- -measure of Z on the boundaries, notice that the intersection of each of the aforementioned hyperplanes $p \cdot x - t = p' \cdot x - t'$ with each boundary $x_i = x_i^{\text{low}}$ is $(n - 2)$ -dimensional, unless the hyperplane coincides with $x_i = x_i^{\text{low}}$. If it is $(n - 2)$ -dimensional then its measure under μ_- is 0. Otherwise, it must be that $p_j = p'_j$, for all $j \neq i$, and $p_i \neq p'_i$; say $p_i > p'_i$ without loss of generality. This implies that (p, t) must dominate (p', t') , for all types $x \in X$. This contradicts our assumption that no menu choices are dominated. \square

Let u be the utility function of the optimal mechanism $\mathcal{M} = (\mathcal{P}, \mathcal{T})$ and γ be the optimal measure of Theorem 2. Then, γ satisfies the properties of Corollary 1. In particular, it holds that:

1.

$$\int u d(\gamma_1 + \mu_-) = \int u d(\mu_+ + \gamma_2) \quad (5.1)$$

2. $u(x) - u(y) = \|x - y\|_1$, $\gamma(x, y)$ almost surely. Since this can happen only if x is coordinate-wise greater than y , it holds (almost surely with respect to γ) that $\|x - y\|_1 = \sum_i x_i - \sum_i y_i$ which implies that (almost surely) $u(x) - \sum_i x_i = u(y) - \sum_i y_i$ and thus

$$\int (u(x) - \sum_i x_i) d\gamma_1 = \int (u(y) - \sum_i y_i) d\gamma_2 \quad (5.2)$$

Moreover, again since x is coordinate-wise greater than y almost surely with respect to γ , it follows that $\gamma_2 \succeq_{cvx(-\bar{1})} \gamma_1$.

We are now ready to use Lemma 9 which follows from Jensen's inequality. We will apply it in two different steps, which we will then combine to show that $\mu_+|_R \preceq_{cvx(\bar{v})} \mu_-|_R$.

Step (ia): We will first apply Lemma 9 to random variables A, B distributed according to the measures $\gamma_2 + \mu_+$ and $\gamma_1 + \mu_-$ respectively. Since $\mu_+ - \mu_- \preceq_{cvx} \gamma_1 - \gamma_2$, by the feasibility of γ , we have that $A \preceq_{cvx} B$. Moreover, $\mathbb{E}[u(A)] = \mathbb{E}[u(B)]$, from Equation (5.1) above, and u is convex and non-decreasing, from the feasibility of u .

To apply Lemma 9, we choose the function $g(x)$, which is a subgradient functions of u , as follows:

- For all $x \in X \setminus Z$ the best choice from the menu of \mathcal{M} is unique, hence the subgradient of u is uniquely defined. For all such x , we set $g(x) = \mathcal{P}(x)$.
- For all other x , u has a continuum of different subgradients at x . In particular, any vector in the convex hull of $\{p : p \cdot x - t = u(x), (p, t) \in \text{Menu}_{\mathcal{M}}\}$ is a valid subgradient. Thus, we can always choose $g(x)$ to equal a vector of probabilities that doesn't appear as an allocation of any choice in menu \mathcal{M} .

Step (ib): it follows from Lemma 9 that there exist random variables $\hat{A} =_{st} A$ and $\hat{B} =_{st} B$ such that, almost surely, $g(\hat{A})$ is a subgradient of u at \hat{B} . Fixing some

$(p, t) \in \text{Menu}_{\mathcal{M}}$ and its corresponding region $R = \{x : p = \mathcal{P}(x)\}$, we denote by $\text{cl}(R) = R \cup \partial R$ the closure of R and by $\text{int}(R) = \text{cl}(R) \setminus Z$ the set of types which strictly prefer (p, t) to any other option in the menu. Note in particular that $\text{int}(R)$ may contain points on the boundary of X . With this notation, we have that almost surely:

$$\hat{B} \in \text{int}(R) \implies \hat{A} \in \text{int}(R); \quad (5.3)$$

$$\hat{A} \in \text{int}(R) \implies \hat{B} \in \text{cl}(R). \quad (5.4)$$

This is because, from Lemma 9, we know that $g(\hat{A})$ is a subgradient of u at \hat{B} almost surely, and we know by definition of $\text{int}(R)$ that the subgradient is unique whenever $\hat{B} \in \text{int}(R)$. Thus, it holds almost surely that whenever $\hat{B} \in \text{int}(R)$ we have $g(\hat{A}) = g(\hat{B})$. Since g is chosen to have differing values on $\text{int}(R)$ and on Z , it follows that whenever $\hat{B} \in \text{int}(R)$, $\hat{A} \in \text{int}(R)$ almost surely. The implication $\hat{A} \in \text{int}(R) \implies \hat{B} \in \text{cl}(R)$ follows from the fact that the subgradient at any point $x \in \text{int}(R)$ can only serve as a subgradient for points $y \in \text{cl}(R)$.

From Lemma 9, we also have that $u(\mathbb{E}[\hat{B}|\hat{A}]) = u(\hat{A})$ almost surely. It follows that, almost surely,

$$u(\mathbb{E}[\hat{B}|\hat{A}]) \cdot \mathbb{I}_{\hat{A} \in \text{int}(R)} = u(\hat{A}) \cdot \mathbb{I}_{\hat{A} \in \text{int}(R)}$$

Given (5.4) and since u is linear restricted to $\text{cl}(R)$, it follows that the left hand side equals:

$$\mathbb{E}[u(\hat{B})|\hat{A}] \cdot \mathbb{I}_{\hat{A} \in \text{int}(R)}$$

We also have from Lemma 9 that, almost surely, it holds componentwise

$$\mathbb{E}[\hat{B}|\hat{A}] \geq \hat{A}. \quad (5.5)$$

The above imply that, almost surely:

$$p_i > 0 \implies \mathbb{E}[\hat{B}_i|\hat{A}] \cdot \mathbb{I}_{\hat{A} \in \text{int}(R)} = \hat{A}_i \cdot \mathbb{I}_{\hat{A} \in \text{int}(R)} \quad (5.6)$$

as otherwise we cannot have $\mathbb{E}[u(\hat{B})|\hat{A}] \cdot \mathbb{I}_{\hat{A} \in \text{int}(R)} = u(\hat{A}) \cdot \mathbb{I}_{\hat{A} \in \text{int}(R)}$, given that u is linear and non-decreasing in $\text{cl}(R)$.

Equations (5.5), (5.6) and Lemma 6 imply that

$$\hat{A} \cdot \mathbb{I}_{\hat{A} \in \text{int}(R)} \preceq_{\text{cvx}(\vec{v})} \hat{B} \cdot \mathbb{I}_{\hat{A} \in \text{int}(R)} \quad (5.7)$$

for the \vec{v} defined in Definition 10 for the menu choice (p, t) . Note that:

$$\begin{aligned} \hat{B} \cdot \mathbb{I}_{\hat{A} \in \text{int}(R)} &= \hat{B} \cdot \mathbb{I}_{\hat{A}, \hat{B} \in \text{int}(R)} + \hat{B} \cdot \mathbb{I}_{\hat{A} \in \text{int}(R) \wedge \hat{B} \notin \text{int}(R)} \\ &= \hat{B} \cdot \mathbb{I}_{\hat{B} \in \text{int}(R)} + \hat{B} \cdot \mathbb{I}_{\hat{A} \in \text{int}(R) \wedge \hat{B} \notin \text{int}(R)} \end{aligned}$$

where for the second equality we used (5.3). Hence, (5.7) implies:

$$\gamma_2|_{\text{int}(R)} + \mu_+|_{\text{int}(R)} \preceq_{\text{cvx}(\vec{v})} \mu_-|_{\text{int}(R)} + \gamma_1|_{\text{int}(R)} + \xi_R \quad (5.8)$$

where ξ_R is the non-negative measure corresponding to $\hat{B} \cdot \mathbb{I}_{\hat{A} \in \text{int}(R) \wedge \hat{B} \notin \text{int}(R)}$ (scaled back appropriately by $\mu_+(X) = \mu_-(X)$).

Step (iia): We will now apply a flipped version of Lemma 9, for convex non-increasing functions,⁸ to the convex function $u(x) - \sum_i x_i$.⁹ We set random variables A', B' distributed according to the measures γ_1 and γ_2 . Since $\gamma_2 \succeq_{\text{cvx}(-\vec{1})} \gamma_1$, we have that $B' \succeq_{\text{cvx}(-\vec{1})} A'$. Moreover, $\mathbb{E}[u(A') - \sum_i A'_i] = \mathbb{E}[u(B') - \sum_i B'_i]$ from Equation (5.2) shown above.

We choose the function $g(x) - \vec{1}$ as the subgradient of $u(x) - \sum_i x_i$.

Step (iib): Fixing any region R and the corresponding $\text{int}(R)$, $\text{cl}(R)$ and \vec{v} as above, we mirror the arguments of Step (i). Now, the version of Lemma 9 for non-increasing functions implies that there exist random variables $\hat{A}' =_{st} A'$ and $\hat{B}' =_{st} B'$ such that,

⁸It is easy to verify that the guarantees of the lemma remain the same except the third guarantee changes to “componentwise smaller than.”

⁹Notice that the partial derivatives are non-positive.

almost surely:

$$\mathbb{E}[\hat{B}'|\hat{A}'] \leq \hat{A}'; \quad (5.9)$$

$$p_i < 1 \implies \mathbb{E}[\hat{B}'_i|\hat{A}'] \cdot \mathbb{I}_{\hat{A}' \in \text{int}(R)} = \hat{A}'_i \cdot \mathbb{I}_{\hat{A}' \in \text{int}(R)}. \quad (5.10)$$

Equations (5.9), (5.10) and Lemma 6 imply that

$$\hat{A}' \cdot \mathbb{I}_{\hat{A}' \in \text{int}(R)} \preceq_{\text{cvx}(\vec{v})} \hat{B}' \cdot \mathbb{I}_{\hat{A}' \in \text{int}(R)} \quad (5.11)$$

and, hence,

$$\gamma_1|_{\text{int}(R)} \preceq_{\text{cvx}(\vec{v})} \gamma_2|_{\text{int}(R)} + \xi'_R, \quad (5.12)$$

where similarly to our derivation above ξ'_R is the non-negative measure corresponding to $\hat{B}' \cdot \mathbb{I}_{\hat{A}' \in \text{int}(R) \wedge \hat{B}' \notin \text{int}(R)}$.

We now combine the results of Steps (i) and (ii) to finish the proof. Combining (5.8) and (5.12), we get that:

$$\mu_+|_{\text{int}(R)} \preceq_{\text{cvx}(\vec{v})} \mu_-|_{\text{int}(R)} + \xi_R + \xi'_R. \quad (5.13)$$

From Proposition 1, it must hold that

$$\mu_+|_{\text{int}(R)}(X) = \mu_-|_{\text{int}(R)}(X) + \xi_R(X) + \xi'_R(X).$$

Summing over all regions and noticing that $\sum_R \mu_-|_{\text{int}(R)}(X) = \mu_-(X)$, from Lemma 10, we get that

$$\mu_+(X) - \mu_+(Z) = \mu_-(X) + \sum_R (\xi_R(X) + \xi'_R(X)).$$

But $\mu_+(X) = \mu_-(X)$, hence $\mu_+(Z) = \sum_R (\xi_R(X) + \xi'_R(X)) = 0$, as all of μ_+ , ξ_R and

ξ'_R are non-negative. Therefore, we can rewrite the property (5.13) as:

$$\mu_+|_R \preceq_{cvx(\vec{v})} \mu_-|_R.$$

This completes the proof.

Chapter 6

A Framework for Computing Optimal Mechanisms

6.1 Constructing Optimal Mechanisms

The results of the previous section characterize optimal mechanisms and give us the tools to check if a mechanism is optimal. In this section, we show how to use the optimal menu conditions we developed to identify candidate mechanisms. In particular, Theorem 3 implies that (in the finite menu case) to find an optimal mechanism we need to identify a set of choices for the menu, such that for every region R that corresponds to a menu outcome it holds that $\mu_+|_R \preceq_{cvx(\vec{v})} \mu_-|_R$ for the appropriate vector \vec{v} . This implies that $\mu_+(R) = \mu_-(R)$, so at the very least the total positive and the total negative mass in each region need to be equal. This property immediately helps us exclude a large class of mechanisms and guides us to identify potential candidates. We note that in this section we will develop techniques which apply not just to finite-menu mechanisms but to mechanisms with infinite menus as well.

We will restrict ourselves to a particularly useful class of mechanisms defined completely by the set of types that are excluded from the mechanism, i.e. they receive no items and pay nothing. We call this set of types the *exclusion set* of a mechanism. The exclusion set gives rise to a mechanism where the utility of a buyer is equal to the ℓ_1 distance between the buyer's type and the closest point in the exclusion set.

All known instances of optimal mechanisms for independently distributed items fall under this category. We proceed to define these concepts formally.

Definition 11 (Exclusion Set). *Let $X = \prod_{i=1}^n [x_i^{low}, x_i^{high}]$. An exclusion set Z of X is a convex, compact, and decreasing¹ subset of X with nonempty interior.*

Definition 12 (Mechanism of an Exclusion Set). *Every exclusion set Z of X induces a mechanism whose utility function $u_Z : X \rightarrow \mathbb{R}$ is defined by:*

$$u_Z(x) = \min_{z \in Z} \|z - x\|_1.$$

Note that, since the exclusion set Z is closed, for any $x \in X$ there exists a $z \in Z$ such that $u_Z(x) = \|z - x\|_1$. Moreover, we show below that any such utility function u_Z satisfies the constraints of the mechanism design problem. That is, the mechanism corresponding to u_Z is IC and IR.

Claim 3. *Let Z be an exclusion set of X . Then u_Z is non-negative, non-decreasing, convex, and has Lipschitz constant (with respect to the ℓ_1 norm) at most 1. In particular, u_Z is the utility function of an incentive compatible and individually rational mechanism.*

Proof. It is obvious that u_Z is non-negative. To show that u_Z is non-decreasing, it suffices to prove that $u_Z(x) \geq u_Z(y)$ for $x, y \in X \setminus Z$ with x component-wise greater than or equal to y . Let $z_x \in Z$ be the closest point to x . Denote by z_y the point with each coordinate being the component-wise minimum of z_x and y . Since Z is decreasing, $z_y \in Z$. We now compute

$$u_Z(x) = \|z_x - x\|_1 = \sum_i |(z_x)_i - x_i| \geq \sum_i |\min\{(z_x)_i, y_i\} - y_i| = \|z_y - y\|_1 \geq u_Z(y)$$

and thus u_Z is non-decreasing.

We will now show that u_Z is convex. Pick arbitrary $x, y \in X$. Denote by z_x and z_y points in Z such that $u_Z(x) = \|x - z_x\|_1$ and $u_Z(y) = \|y - z_y\|_1$. Since Z is convex,

¹A decreasing subset $Z \subset X$ satisfies the property that for all $a, b \in X$ such that a is component-wise less than or equal to b , if $b \in Z$ then $a \in Z$ as well.

the point $(z_x + z_y)/2$ is in Z . Thus

$$u_Z\left(\frac{x+y}{2}\right) \leq \left\| \frac{x+y}{2} - \frac{z_x+z_y}{2} \right\|_1 \leq \frac{\|x-z_x\|_1 + \|y-z_y\|_1}{2} = \frac{u_Z(x) + u_Z(y)}{2}$$

and therefore u_Z is convex.

Lastly, we verify that u_Z has Lipschitz constant at most 1. Indeed,

$$u_Z(x) - u_Z(y) \leq \|x - z_y\|_1 - u_Z(y) = \|x - z_y\|_1 - \|y - z_y\|_1 \leq \|x - y\|_1.$$

□

6.1.1 Constructing Optimal Mechanisms for 2 Items

To provide sufficient conditions for u_Z to be optimal for the case of 2 items, we define the concept of a canonical partition. A canonical partition divides X into regions such that the mechanism's allocation function within each region has a similar form. Roughly, the canonical partition separates X based on which direction (either “down,” “left,” or “diagonally”) one must travel to reach the closest point in Z . While the definition is involved, the geometric picture of Figure 6-1 is straightforward.

Definition 13 (Critical price, Critical point, Outer boundary functions). *Let Z be an exclusion set of X . Denote by P the maximum value $P = \max\{x+y : (x, y) \in Z\}$, we call P the critical price. We now define the critical point (x_{crit}, y_{crit}) , such that*

$$x_{crit} = \min\{x : (x, P-x) \in Z\} \text{ and } y_{crit} = \min\{y : (P-y, y) \in Z\}$$

We define the outer boundary functions of Z to be the functions s_1, s_2 given by

$$s_1(x) = \max\{y : (x, y) \in Z\} \text{ and } s_2(y) = \max\{x : (x, y) \in Z\},$$

with domain $[0, x_{crit}]$ and $[0, y_{crit}]$ respectively.

Definition 14 (Canonical partition). *Let Z be an exclusion set of X with critical point (x_{crit}, y_{crit}) as in Definition 13. We define the canonical partition of X induced*

by Z to be the partition of X into $Z \cup \mathcal{A} \cup \mathcal{B} \cup \mathcal{W}$, where

$$\mathcal{A} = \{(x, y) \in X : x < x_{crit}\} \setminus Z; \quad \mathcal{B} = \{(x, y) \in X : y < y_{crit}\} \setminus Z; \quad \mathcal{W} = X \setminus (Z \cup \mathcal{A} \cup \mathcal{B}),$$

as shown in Figure 6-1.

Note that the outer boundary functions s_1, s_2 of an exclusion set Z are concave and thus are differentiable almost everywhere on $[0, c_1]$ and have non-increasing derivatives.

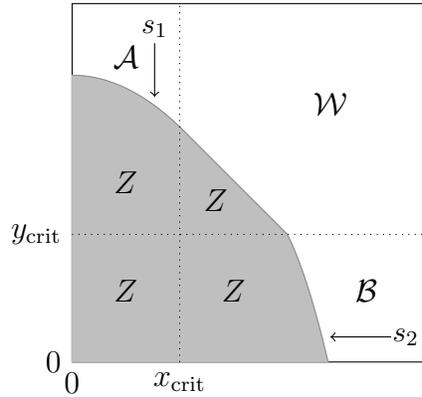


Figure 6-1: The canonical partition

We now restate the utility function u_Z of a mechanism with exclusion set Z in terms of a canonical partition.

Claim 4. *Let Z be an exclusion set of X with outer boundary functions s_1, s_2 and critical price P , and let $Z \cup \mathcal{A} \cup \mathcal{B} \cup \mathcal{W}$ be its canonical partition. Then for all $(v_1, v_2) \in X$, the utility function u_Z of the mechanism with exclusion set Z is given by:*

$$u_Z(v_1, v_2) = \begin{cases} 0 & \text{if } (v_1, v_2) \in Z \\ v_2 - s_1(v_1) & \text{if } (v_1, v_2) \in \mathcal{A} \\ v_1 - s_2(v_2) & \text{if } (v_1, v_2) \in \mathcal{B} \\ v_1 + v_2 - P & \text{if } (v_1, v_2) \in \mathcal{W}. \end{cases}$$

Proof. The proof is fairly straightforward casework. We prove one of the cases here, and the remaining cases are similar.

Pick any $v = (v_1, v_2) \in \mathcal{A}$. We will show that the closest $z \in Z$ is the point $z^* = (v_1, s_1(v_1))$. Pick $z' = (z'_1, z'_2) \in Z$ such that $u_Z(v) = \|v - z'\|_1$. It must be the case that $z'_1 \leq v_1$, since otherwise (v_1, z'_2) would be in Z (as Z is decreasing) and strictly closer to v .

We now have that $\|v - z'\|_1 \geq \|v\|_1 - \|z'\|_1 \geq \|v\|_1 - \max_{x \in [0, v_1]}(x + s_1(x))$. Since the less restricted maximization problem, $\max_{x \in [0, x_{\text{crit}}]}(x + s_1(x))$ is maximized at x_{crit} and the function $(x + s_1(x))$ is concave, the maximum of the more constrained version is achieved at $x = v_1$. Thus, we have that, $\|v - z'\|_1 \geq \|v\|_1 - v_1 - s_1(v_1) = v_2 - s_1(v_1) = \|v - z^*\|_1$. \square

We now describe sufficient conditions under which u_Z is optimal.

Definition 15 (Well-formed canonical partition). *Let $Z \cup \mathcal{A} \cup \mathcal{B} \cup \mathcal{W}$ be a canonical partition of X induced by exclusion set Z and let μ be a signed Radon measure on X such that $\mu(X) = 0$. We say that the canonical partition is well-formed with respect to μ if the following conditions are satisfied:*

1. $\mu|_Z \preceq_{\text{cvx}} 0$ and $\mu|_{\mathcal{W}} \succeq_2 0$, and
2. for all $v \in X$ and all $\epsilon > 0$:
 - $\mu|_{\mathcal{A}}([v_1, v_1 + \epsilon] \times [v_2, \infty)) \geq 0$, with equality whenever $v_2 = 0$
 - $\mu|_{\mathcal{B}}([v_1, \infty) \times [v_2, v_2 + \epsilon]) \geq 0$, with equality whenever $v_1 = 0$

We point out the similarities between a well-formed canonical partition and the sufficient conditions for menu optimality of Theorem 3. Condition 1 gives exactly the stochastic dominance conditions that need to hold in regions Z and \mathcal{W} . We interpret Condition 2 as saying that $\mu|_{\mathcal{A}}$ (resp. $\mu|_{\mathcal{B}}$) allows for the positive mass in any vertical (resp. horizontal) “strip” to be matched to the negative mass in the strip by only transporting “downwards” (resp. “leftwards”). These conditions, guarantee (single-dimensional) first order dominance of the measures along each strip which is stronger requirement than the convex dominance conditions of Theorem 3. In practice, when μ is given by a density function, we verify these conditions by analyzing the integral

of the density function along appropriate vertical or horizontal lines. Even though Theorem 3 applies only for mechanisms with finite menus, we prove in Theorem 7 that a mechanism induced by an exclusion set is optimal for a 2-item instance if the canonical partition of its exclusion set is well-formed. Refer back to Figure 6-1 to visualize such a mechanism.

Theorem 7. *Let μ be the transformed measure of a probability density function f . If there exists an exclusion set Z inducing a canonical partition $Z \cup \mathcal{A} \cup \mathcal{B} \cup \mathcal{W}$ of X that is well-formed with respect to μ , then the optimal IC and IR mechanism for a single additive buyer whose values for two goods are distributed according to the joint distribution f is the mechanism induced by exclusion set Z . In particular, the mechanism uses the following allocation and price for a buyer with reported type $(x, y) \in X$:*

- if $(x, y) \in Z$, the buyer receives no goods and is charged 0;
- if $(x, y) \in \mathcal{A}$, the buyer receives item 1 with probability $-s'_1(x)$, item 2 with probability 1, and is charged $s_1(x) - xs'_1(x)$;
- if $(x, y) \in \mathcal{B}$, the buyer receives item 2 with probability $-s'_2(y)$, item 1 with probability 1, and is charged $s_2(y) - ys'_2(y)$;
- if $(x, y) \in \mathcal{W}$, the buyer receives both goods with probability 1 and is charged P ;

where s_1, s_2 are the boundary functions and P is the critical price as in Definition 13.

Proof. We will show that u_Z maximizes $\sup_{u \in \mathcal{U}(X) \cap \mathcal{L}_1(X)} \int_X u d\mu$. By Corollary 1, it suffices to provide a $\gamma \in \Gamma_+(X \times X)$ such that $\gamma_1 - \gamma_2 \succeq_{cvx} \mu$, $\int u_Z d(\gamma_1 - \gamma_2) = \int u_Z d\mu$, and $u_Z(x) - u_Z(y) = \|x - y\|_1$ holds γ -almost surely. The γ we construct will never transport mass between regions. That is, $\gamma = \gamma_Z + \gamma_W + \gamma_A + \gamma_B$ where²

- $\gamma_Z = 0$. We notice that $(\gamma_Z)_1 - (\gamma_Z)_2 = 0 \succeq_{cvx} \mu|_Z$.

²We chose this notation for simplicity, where $\gamma_Z \in \Gamma_+(Z \times Z)$, $\gamma_W \in \Gamma_+(W \times W)$, and so on.

- $\gamma_{\mathcal{W}}$ is constructed such that $(\gamma_{\mathcal{W}})_1 - (\gamma_{\mathcal{W}})_2 \succeq_{cvx} \mu|_{\mathcal{W}}$ and the component-wise inequality $x \geq y$ holds $\gamma_{\mathcal{W}}(x, y)$ almost surely.³ As in our proof of Theorem 3, the existence of such a $\gamma_{\mathcal{W}}$ is guaranteed by Strassen’s theorem for second order dominance (presented in Lemma 5).
- $\gamma_{\mathcal{A}} \in \Gamma_+(\mathcal{A} \times \mathcal{A})$ will be constructed to have respective marginals $\mu_+|_{\mathcal{A}}$ and $\mu_-|_{\mathcal{A}}$, and so that, $\gamma_{\mathcal{A}}(x, y)$ almost surely, it holds that $x_1 = y_1$ and $x_2 \geq y_2$. Thus, $(\gamma_{\mathcal{A}})_1 - (\gamma_{\mathcal{A}})_2 = \mu|_{\mathcal{A}}$, and $\gamma_{\mathcal{A}}$ sends positive mass “downwards.”⁴ We claim that such a map can indeed be constructed, by noticing that Property 2 of Definition 15 guarantees that, restricted to any vertical strip inside \mathcal{A} , μ_+ first-order stochastically dominates μ_- .⁵ Hence, Strassen’s theorem for first-order dominance guarantees that restricted to that strip μ_+ can be coupled with μ_- so that, with probability 1, mass is only moved downwards.

Measure $\gamma_{\mathcal{A}}$ satisfies $x_1 = y_1$, $\gamma_{\mathcal{A}}(x, y)$ almost surely, and hence also

$$u_Z(x) - u_Z(y) = (x_2 - s(x_1)) - (y_2 - s(y_1)) = x_2 - y_2 = \|x - y\|_1.$$

- $\gamma_{\mathcal{B}} \in \Gamma_+(\mathcal{B} \times \mathcal{B})$ is constructed analogously to $\gamma_{\mathcal{A}}$, except sending mass “leftwards.” That is, $\gamma_{\mathcal{B}}(x, y)$ almost-surely, the relationships $x_1 \geq y_1$ and $x_2 = y_2$ hold.

It follows by our construction that $\gamma = \gamma_Z + \gamma_{\mathcal{W}} + \gamma_{\mathcal{A}} + \gamma_{\mathcal{B}}$ satisfies all necessary properties to certify optimality of u_Z . \square

6.2 Verifying Stochastic Dominance

A technical difficulty is verifying the stochastic dominance relation $\mu|_{\mathcal{W}} \succeq_2 0$ required to apply the theorem. In our examples, we will have the stronger condition $\mu|_{\mathcal{W}} \succeq_1 0$,

³As in Example 2 and as discussed in Remark 1, we aim for $\gamma_{\mathcal{W}}$ to transport “downwards and leftwards” since both items are allocated with probability 1 in \mathcal{W} .

⁴Once again, the intuition for this construction follows Remark 1.

⁵Indeed, as $\epsilon \rightarrow 0$, Property 2 states exactly the one-dimensional equivalent condition for first-order stochastic dominance in terms of cumulative density functions.

which is easier to verify, yet still imposes technical difficulties. We present a useful tool, Lemma 11, for verifying first-order stochastic dominance. In Section 6.3 and Section 6.4, we then provide example applications of Theorem 7 and Lemma 11 to solve for optimal mechanisms.

A useful tool for verifying first order dominance between measures is the following.

Lemma 11. *Let $\mathcal{C} = [p_1, q_1) \times [p_2, q_2)$ where q_1 and q_2 are possibly infinite and let R be a decreasing nonempty subset of \mathcal{C} . Consider two measures $\kappa, \lambda \in \Gamma_+(\mathcal{C})$ with bounded integrable density functions $g, h : \mathcal{C} \rightarrow \mathbb{R}_{\geq 0}$ respectively that satisfy the conditions:*

- $g(x, y) = h(x, y) = 0$ for all $(x, y) \in R$.
- $\int_{\mathcal{C}} g(x, y) dx dy = \int_{\mathcal{C}} h(x, y) dx dy$.
- For any basis vector $e_i \in \{e_1 \equiv (1, 0), e_2 \equiv (0, 1)\}$ and any point $z \in R$:

$$\int_0^{q_i - z_i} g(z + \tau e_i) - h(z + \tau e_i) d\tau \leq 0.$$

- There exist non-negative functions $\alpha : [p_1, q_1) \rightarrow \mathbb{R}_{\geq 0}$ and $\beta : [p_2, q_2) \rightarrow \mathbb{R}_{\geq 0}$, and an increasing function $\eta : \mathcal{C} \rightarrow \mathbb{R}$ such that for all $(x, y) \in \mathcal{C} \setminus R$:

$$g(x, y) - h(x, y) = \alpha(x) \cdot \beta(y) \cdot \eta(x, y)$$

Then $\kappa \succeq_1 \lambda$.

Lemma 11 provides a sufficient condition for a measure to stochastically dominate another in the first order. Its proof is given in Section 6.2.1 and is an application of a claim which states that an equivalent condition for first-order stochastic dominance is that one measure has more mass than the other on all sets that are unions of *finitely many* “increasing boxes.” When the conditions of Lemma 11 are satisfied, we can induct on the number of boxes by removing one box at a time. We note that Lemma 11 is applicable even to distributions with unbounded support.

Interpreting the Conditions of Lemma 11: Lemma 11 is applicable whenever two density functions, g and h , are nonzero on some set $\mathcal{C} \setminus R$, where R is a decreasing subset of some two-dimensional box \mathcal{C} . This setting is motivated by Figure 6-1 and Theorem 7. Recall that, in order to apply Theorem 7, we need to check a second order stochastic dominance condition in region \mathcal{W} , namely $\mu|_{\mathcal{W}} \succeq_2 0$.

While Theorem 7 demands checking a second order stochastic dominance condition, an easier and sufficient goal is to check first order stochastic dominance, namely $\mu|_{\mathcal{W}} \succeq_1 0$. To do this, we can readily use Lemma 11, by taking $\mathcal{C} = [x_{\text{crit}}, \infty) \times [y_{\text{crit}}, \infty)$, $R = \mathcal{C} \cap Z$, and g, h the densities corresponding to measures $\mu_+|_{\mathcal{W}}$ and $\mu_-|_{\mathcal{W}}$. The way region \mathcal{W} is defined in Theorem 7 guarantees that the two measures have equal mass, so the first two conditions of the lemma will be satisfied automatically. For the third condition, we need to verify that, if we integrate $g - h$ along either a vertical or a horizontal line outwards starting from any point in R , the result is *non-positive*. The last condition of Lemma 11 requires that the density function of the measure $\mu|_{\mathcal{W}}$, i.e. $g - h$, have an appropriate form. If the values of the buyer for the two items are independently distributed according to distributions with densities f_1 and f_2 , then the density of measure μ in the interior according to Equation 3.3 can be written as $-f_1(x)f_2(y) \left(\frac{f_1'(x)x}{f_1(x)} + \frac{f_2'(y)y}{f_2(y)} + 3 \right)$. The last condition of the lemma is thus satisfied if the functions $\frac{f_1'(x)x}{f_1(x)}$ and $\frac{f_2'(y)y}{f_2(y)}$ are decreasing, a condition that is easy to verify.

6.2.1 Proof of Lemma 11

We begin with the standard result that a sufficient condition for first-order stochastic dominance is that one measure assigns more mass than the other to all increasing sets.

Claim 5. *Let α, β be positive finite Radon measures on $\mathbb{R}_{\geq 0}^n$ with $\alpha(\mathbb{R}_{\geq 0}^n) = \beta(\mathbb{R}_{\geq 0}^n)$.*

A necessary and sufficient condition for $\alpha \succeq_1 \beta$ is that for all increasing⁶ measurable sets A , $\alpha(A) \geq \beta(A)$.

PROOF OF CLAIM 5: Without loss of generality assume that $\alpha(\mathbb{R}_{\geq 0}^n) = \beta(\mathbb{R}_{\geq 0}^n) = 1$.

It is obvious that the condition is necessary by considering the indicator function of any increasing set A . To prove sufficiency, suppose that the condition holds and that on the contrary, α does not stochastically dominate β . Then there exists an increasing, bounded, measurable function f such that

$$\int f d\beta - \int f d\alpha > 2^{-k+1}$$

for some positive integer k . Without loss of generality, we may assume that f is nonnegative, by adding the constant of $-f(0)$ to all values. We now define the function \tilde{f} by point-wise rounding f upwards to the nearest multiple of 2^{-k} . Clearly \tilde{f} is increasing, measurable, and bounded. Furthermore, we have

$$\int \tilde{f} d\beta - \int \tilde{f} d\alpha \geq \int f d\beta - \int f d\alpha - 2^{-k} > 2^{-k+1} - 2^{-k} > 0.$$

We notice, however, that \tilde{f} can be decomposed into the weighted sum of indicator functions of increasing sets. Indeed, let $\{r_1, \dots, r_m\}$ be the set of all values taken by \tilde{f} , where $r_1 > r_2 > \dots > r_m$. We notice that, for any $s \in \{1, \dots, m\}$, the set $A_s = \{z : \tilde{f}(z) \geq r_s\}$ is increasing and measurable. Therefore, we may write

$$\tilde{f} = \sum_{s=1}^m (r_s - r_{s-1}) I_s$$

where I_s is the indicator function for A_s and where we set $r_0 = 0$. We now compute

$$\int \tilde{f} d\beta = \sum_{s=1}^m (r_s - r_{s-1}) \beta(A_s) \leq \sum_{s=1}^m (r_s - r_{s-1}) \alpha(A_s) = \int \tilde{f} d\alpha,$$

contradicting the fact that $\int \tilde{f} d\beta > \int \tilde{f} d\alpha$. □

⁶An increasing set $A \subset \mathbb{R}_{\geq 0}^n$ satisfies the property that for all $a, b \in \mathbb{R}_{\geq 0}^n$ such that a is component-wise greater than or equal to b , if $b \in A$ then $a \in A$ as well.

Due to Claim 5, to verify that a measure α stochastically dominates β in the first order, we must ensure that $\alpha(A) \geq \beta(A)$ for all increasing measurable sets A . This verification might still be difficult, since an increasing set can have fairly unconstrained structure. In Lemma 13 we simplify this task by showing that we need not verify the inequality for all increasing A , but rather only for a special class of increasing subsets.

Definition 16. For any $z \in \mathbb{R}_{\geq 0}^n$, we define the base rooted at z to be

$$B_z \triangleq \{z' : z \preceq z'\},$$

the minimal increasing set containing z , where the notation $z \preceq z'$ denotes that every component of z is at most the corresponding component of z' .

We denote by Q_k to be the set of points in $\mathbb{R}_{\geq 0}^n$ with all coordinates multiples of 2^{-k} .

Definition 17. An increasing set S is k -discretized if $S = \bigcup_{z \in S \cap Q_k} B_z$. A corner c of a k -discretized set S is a point $c \in S \cap Q_k$ such that there does not exist $z \in S \setminus \{c\}$ with $z \preceq c$.

Lemma 12. Every k -discretized set S has only finitely many corners. Furthermore, $S = \bigcup_{c \in \mathcal{C}} B_c$, where \mathcal{C} is the collection of corners of S .

PROOF OF LEMMA 12: We prove that there are finitely many corners by induction on the dimension, n . In the case $n = 1$ the result is obvious, since if S is nonempty it has exactly one corner. Now suppose S has dimension n . Pick some corner $\hat{c} = (c_1, \dots, c_n) \in S$. We know that any other corner must be strictly less than \hat{c} in some coordinate. Therefore,

$$|\mathcal{C}| \leq 1 + \sum_{i=1}^n |\{c \in \mathcal{C} \text{ s.t. } c_i < \hat{c}_i\}| = 1 + \sum_{i=1}^n \sum_{j=1}^{2^k \hat{c}_i} |\{c \in \mathcal{C} \text{ s.t. } c_i = \hat{c}_i - 2^{-k} j\}|.$$

By the inductive hypothesis, we know that each set $\{c \in \mathcal{C} \text{ s.t. } c_i = \hat{c}_i - 2^{-k} j\}$ is

finite, since it is contained in the set of corners of the $(n - 1)$ -dimensional subset of S whose points have i^{th} coordinate $\hat{c}_i - 2^{-k}j$. Therefore, $|\mathcal{C}|$ is finite.

To show that $S = \bigcup_{c \in \mathcal{C}} B_c$, pick any $z \in S$. Since S is k -discretized, there exists a $b \in S \cap Q_k$ such that $z \in B_b$. If b is a corner, then z is clearly contained in $\bigcup_{c \in \mathcal{C}} B_c$. If b is not a corner, then there is some other point $b' \in S \cap Q_k$ with $b' \preceq b$. If b' is a corner, we're done. Otherwise, we repeat this process at most $2^k \sum_j b_j$ times, after which time we will have reached a corner c of S . By construction, we have $z \in B_c$, as desired. \square

We now show that, to verify that one measure dominates another on all increasing sets, it suffices to verify that this holds for all sets that are the union of finitely many bases.

Lemma 13. *Let $g, h : \mathbb{R}_{\geq 0}^n \rightarrow \mathbb{R}_{\geq 0}$ be bounded integrable functions such that $\int_{\mathbb{R}_{\geq 0}^n} g(x)dx$ and $\int_{\mathbb{R}_{\geq 0}^n} h(x)dx$ are finite. Suppose that, for all finite collections Z of points in $\mathbb{R}_{\geq 0}^n$, we have*

$$\int_{\bigcup_{z \in Z} B_z} g(x)dx \geq \int_{\bigcup_{z \in Z} B_z} h(x)dx.$$

Then for all increasing sets $A \subseteq \mathbb{R}_{\geq 0}^n$,

$$\int_A g(x)dx \geq \int_A h(x)dx.$$

PROOF OF LEMMA 13: Let A be an increasing set. We clearly have $A = \bigcup_{z \in A} B_z$. For any point $z \in \mathbb{R}_{\geq 0}^n$, denote by $z^{n,k}$ the point in $\mathbb{R}_{\geq 0}^n$ such that for each component i , the i^{th} component of $z^{n,k}$ is the maximum of 0 and $z_i - 2^{-k}$.

We define the following two sets, which we think of as approximations of A :

$$A_k^l \triangleq \bigcup_{z \in A \cap Q_k} B_z; \quad A_k^u \triangleq \bigcup_{z \in A \cap Q_k} B_{z^{n,k}}.$$

It is clear that both A_k^l and A_k^u are k -discretized. Furthermore, for any $z \in A$ there exists a $z' \in A \cap Q_k$ such that each component of z' is at most 2^{-k} more than the corresponding component of z . Therefore $A_k^l \subseteq A \subseteq A_k^u$.

We now will bound

$$\int_{A_k^u} g(x)dx - \int_{A_k^l} g(x)dx.$$

Let

$$W_k = \{z \in \mathbb{R}_{\geq 0}^n : z_i > k \text{ for some } i\}; \quad W_k^c = \{z \in \mathbb{R}_{\geq 0}^n : z_i \leq k \text{ for all } i\}.$$

The set W_k^c contains all points which are lie inside in a box of side length k rooted at the origin, and W_k contains all points outside of this box. We have the immediate (loose) bound that

$$\int_{A_k^u \cap W_k} gdx - \int_{A_k^l \cap W_k} gdx \leq \int_{W_k} gdx.$$

Furthermore, since $\lim_{k \rightarrow \infty} \int_{W_k^c} gdx = \int_{\mathbb{R}_{\geq 0}^n} gdx$, we know that $\lim_{k \rightarrow \infty} \int_{W_k} gdx = 0$.

Therefore,

$$\lim_{k \rightarrow \infty} \left(\int_{A_k^u \cap W_k} gdx - \int_{A_k^l \cap W_k} gdx \right) = 0.$$

Next, we bound

$$\int_{A_k^u \cap W_k^c} gdx - \int_{A_k^l \cap W_k^c} gdx \leq |g|_{\sup} (V(A_k^u \cap W_k^c) - V(A_k^l \cap W_k^c))$$

where $|g|_{\sup} < \infty$ is the supremum of g , and $V(\cdot)$ denotes the Lebesgue measure.

For each $m \in \{1, \dots, n+1\}$ and $z \in \mathbb{R}_{\geq 0}^n$, we define the point $z^{m,k}$ by:

$$z_i^{m,k} = \begin{cases} \max\{0, z_i - 2^{-k}\} & \text{if } i < m \\ z_i & \text{otherwise} \end{cases}$$

and set

$$A_k^m \triangleq \bigcup_{z \in A \cap Q_k} B_{z^{m,k}}.$$

We have, by construction, $A_k^l = A_k^1$ and $A_k^u = A_k^{n+1}$. Therefore,

$$V(A_k^u \cap W_k^c) - V(A_k^l \cap W_k^c) = \sum_{m=1}^n (V(A_k^{m+1} \cap W_k^c) - V(A_k^m \cap W_k^c)).$$

We notice that, for any point $(z_1, z_2, \dots, z_{m-1}, z_{m+1}, \dots, z_n) \in [0, k]^{n-1}$, there is an interval I of length at most 2^{-k} such that

$$(z_1, z_2, \dots, z_{m-1}, w, z_{m+1}, \dots, z_n) \in (A_k^{m+1} \setminus A_k^m) \cap W_k^c$$

if and only if $w \in I$. Therefore,

$$\begin{aligned} V(A_k^{m+1} \cap W_k^c) - V(A_k^m \cap W_k^c) \\ \leq \int_0^k \cdots \int_0^k \int_0^k \cdots \int_0^k 2^{-k} dz_1 \cdots dz_{m-1} dz_{m+1} \cdots dz_n = 2^{-k} k^{n-1}. \end{aligned}$$

We thus have the bound

$$|g|_{\sup} (V(A_k^u \cap W_k^c) - V(A_k^l \cap W_k^c)) \leq |g|_{\sup} \sum_{m=1}^n 2^{-k} k^{n-1} = n|g|_{\sup} 2^{-k} k^{n-1}$$

and therefore

$$\begin{aligned} \int_{A_k^u} g dx - \int_{A_k^l} g dx &= \int_{A_k^u \cap W_k} g dx - \int_{A_k^l \cap W_k} g dx + \int_{A_k^u \cap W_k^c} g dx - \int_{A_k^l \cap W_k^c} g dx \\ &\leq \left(\int_{A_k^u \cap W_k} g dx - \int_{A_k^l \cap W_k} g dx \right) + n|g|_{\sup} 2^{-k} k^{n-1}. \end{aligned}$$

In particular, we have

$$\lim_{k \rightarrow \infty} \left(\int_{A_k^u} g dx - \int_{A_k^l} g dx \right) = 0.$$

Since $\int_{A_k^u} g dx \geq \int_A g dx \geq \int_{A_k^l} g dx$, we have

$$\lim_{k \rightarrow \infty} \int_{A_k^u} g dx = \int_A g dx = \lim_{k \rightarrow \infty} \int_{A_k^l} g dx.$$

Similarly, we have

$$\int_A h dx = \lim_{k \rightarrow \infty} \int_{A_k^l} h dx$$

and thus

$$\int_A (g - h) dx = \lim_{k \rightarrow \infty} \left(\int_{A_k^l} g dx - \int_{A_k^l} h dx \right).$$

Since A_k^l is k -discretized, it has finitely many corners. Letting Z_k denote the corners of A_k^l , we have $A_k^l = \bigcup_{z \in Z_k} B_z$, and thus by our assumption $\int_{A_k^l} g dx - \int_{A_k^l} h dx \geq 0$ for all k . Therefore $\int_A (g - h) dx \geq 0$, as desired. \square

We are now ready to prove Lemma 11.

PROOF OF LEMMA 11:

We begin by defining, for any a and b with $p_1 \leq a \leq b \leq q_1$, the function $\zeta_a^b : [p_2, q_2] \rightarrow \mathbb{R}$ by

$$\zeta_a^b(w_2) \triangleq \int_a^b (g(z_1, w_2) - h(z_1, w_2)) dz_1.$$

This function $\zeta_a^b(w_2)$ represents the integral of $g - h$ along the vertical line from (a, w_2) to (b, w_2) .

Claim 6. *If $(a, w_2) \in R$, then $\zeta_a^b(w_2) \leq 0$.*

PROOF OF CLAIM 6: The inequality trivially holds unless there exists a $z_1 \in [a, b]$ such that $g(z_1, w_2) > h(z_1, w_2)$, so suppose such a z_1 exists. It must be that $(z_1, w_2) \notin R$, since both g and h are 0 in R . Indeed, because R is a decreasing set it is also true that $(\tilde{z}_1, w_2) \notin R$ for all $\tilde{z}_1 \geq z_1$. This implies by our assumption that

$$g(\tilde{z}_1, w_2) - h(\tilde{z}_1, w_2) = \alpha(\tilde{z}_1) \cdot \beta(w_2) \cdot \eta(\tilde{z}_1, w_2),$$

for all $\tilde{z}_1 \geq z_1$. Given that $g(z_1, w_2) > h(z_1, w_2)$ and that $\eta(\cdot, w_2)$ is an increasing function, we know that $g(\tilde{z}_1, w_2) \geq h(\tilde{z}_1, w_2)$ for all $\tilde{z}_1 \geq z_1$. Therefore, we have

$$\zeta_a^{z_1}(w_2) \leq \zeta_a^b(w_2) \leq \zeta_a^{q_1}(w_2).$$

We notice, however, that $\zeta_a^{q_1}(w_2) \leq 0$ by assumption, and thus the claim is proven. \square

We now claim the following:

Claim 7. *Suppose that $\zeta_a^b(w_2^*) > 0$ for some $w_2^* \in [c_2, q_2)$. Then $\zeta_a^b(w_2) \geq 0$ for all $w_2 \in [w_2^*, q_2)$.*

PROOF OF CLAIM 7: Given that $\zeta_a^b(w_2^*) > 0$, our previous claim implies that $(a, w_2^*) \notin R$. Furthermore, since R is a decreasing set and $w_2 \geq w_2^*$, follows that $(a, w_2) \notin R$, and furthermore that $(c, w_2) \notin R$ for any $c \geq a$ in $[c_1, q_1)$. Therefore, we may write

$$\zeta_a^b(w_2) = \int_a^b (g(z_1, w_2) - h(z_1, w_2)) dz_1 = \int_a^b (\alpha(z_1) \cdot \beta(w_2) \cdot \eta(z_1, w_2)) dz_1.$$

Similarly, $(c, w_2^*) \notin R$ for any $c \geq a$, so

$$\zeta_a^b(w_2^*) = \int_a^b (\alpha(z_1) \cdot \beta(w_2^*) \cdot \eta(z_1, w_2^*)) dz_1.$$

Note that, since $\zeta_a^b(w_2^*) > 0$, we have $\beta(w_2^*) > 0$. Thus, since η is increasing,

$$\zeta_a^b(w_2) \geq \int_a^b (\alpha(z_1) \cdot \beta(w_2) \cdot \eta(z_1, w_2^*)) dz_1 = \frac{\beta(w_2)}{\beta(w_2^*)} \zeta_a^b(w_2^*) \geq 0,$$

as desired. □

We extend g and h to all of $\mathbb{R}_{\geq 0}^2$ by setting them to be 0 outside of \mathcal{C} . By Claim 13, to prove that $g \succeq_1 h$ it suffices to prove that $\int_A g dx dy \geq \int_A h dx dy$ for all sets A which are the union of finitely many bases. Since g and h are 0 outside of \mathcal{C} , it suffices to consider only bases $B_{z'}$ where $z' \in \mathcal{C}$, since otherwise we can either remove the base (if it is disjoint from \mathcal{C}) or can increase the coordinates of z' moving it to \mathcal{C} without affecting the value of either integral.

We now complete the proof of Lemma 11 by induction on the number of bases in the union.

Base Case. We aim to show $\int_{B_r} (g - h) dx dy \geq 0$ for any $r = (r_1, r_2) \in \mathcal{C}$. We have

$$\int_{B_r} (g - h) dx dy = \int_{r_2}^{q_2} \int_{r_1}^{q_1} (g - h) dz_1 dz_2 = \int_{r_2}^{q_2} \zeta_{r_1}^{q_1}(z_2) dz_2.$$

By Claim 7, we know that either $\zeta_{r_1}^{q_1}(z_2) \geq 0$ for all $z_2 \geq r_2$, or $\zeta_{r_1}^{q_1}(z_2) \leq 0$ for all z_2 between p_2 and r_2 . In the first case, the integral is clearly nonnegative, so we may

assume that we are in the second case. We then have

$$\begin{aligned} \int_{r_2}^{q_2} \zeta_{r_1}^{q_1}(z_2) dz_2 &\geq \int_{p_2}^{q_2} \zeta_{r_1}^{q_1}(z_2) dz_2 = \int_{p_2}^{q_2} \int_{r_1}^{q_1} (g-h) dz_1 dz_2 \\ &= \int_{r_1}^{q_1} \int_{p_2}^{q_2} (g-h) dz_2 dz_1. \end{aligned}$$

By an analogous argument to that above, we know that either $\int_{p_2}^{q_2} (g-h)(z_1, z_2) dz_2$ is nonnegative for all $z_1 \geq r_1$ (in which case the desired inequality holds trivially) or is nonpositive for all z_1 between p_1 and r_1 . We assume therefore that we are in the second case, and thus

$$\int_{r_1}^{q_1} \int_{p_2}^{q_2} (g-h) dz_2 dz_1 \geq \int_{p_1}^{q_1} \int_{p_2}^{q_2} (g-h) dz_2 dz_1 = \int_{\mathcal{C}} (g-h) dx dy,$$

which is nonnegative by assumption.

Inductive Step. Suppose that we have proven the result for all sets which are finite unions of at most k bases. Consider now a set

$$A = \bigcup_{i=1}^{k+1} B_{z^{(i)}}.$$

We may assume that all $z^{(i)}$ are distinct and that there do not exist distinct $z^{(i)}, z^{(j)}$ with $z^{(i)}$ component-wise less than $z^{(j)}$, since otherwise we could remove one such $B_{z^{(i)}}$ from the union without affecting the set A and the desired inequality would follow from the inductive hypothesis.

We may therefore order the $z^{(i)}$ such that

$$p_1 \leq z_1^{(k+1)} < z_1^{(k)} < z_1^{(k-1)} < \dots < z_1^{(1)}$$

$$p_2 \leq z_2^{(1)} < z_2^{(2)} < z_2^{(3)} < \dots < z_2^{(k+1)}.$$

By Claim 7, we know that one of the two following cases must hold:

Case 1: $\zeta_{z_1^{(k+1)}}^{z_1^{(k)}}(w_2) \leq 0$ for all $p_2 \leq w_2 \leq z_2^{(k+1)}$.

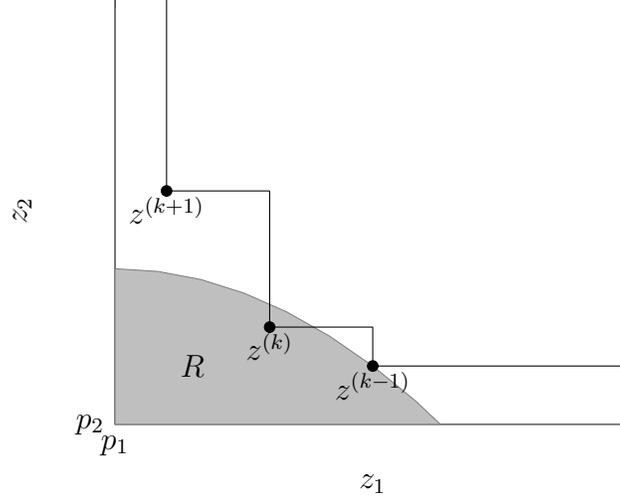


Figure 6-2: We show that either decreasing $z_2^{(k+1)}$ to $z_2^{(k)}$ or removing $z^{(k+1)}$ entirely decreases the value of $\int_A (f - g)$. In either case, we can apply our inductive hypothesis.

In this case, we see that

$$\int_{z_2^{(k)}}^{z_2^{(k+1)}} \int_{z_1^{(k+1)}}^{z_1^{(k)}} (f - g) dz_1 dz_2 = \int_{z_2^{(k)}}^{z_2^{(k+1)}} \zeta_{z_1^{(k+1)}}^{z_1^{(k)}}(w) dw \leq 0.$$

For notational purposes, we denote here by $(f - g)(S)$ the integral $\int_S (f - g) dz_1 dz_2$ for any set S . We compute

$$\begin{aligned} (f - g)(A) &\geq (f - g)(A) \\ &\quad + (f - g) \left(\left\{ z : z_1^{(k+1)} \leq z_1 \leq z_1^{(k)} \text{ and } z_2^{(k)} \leq z_2 \leq z_2^{(k+1)} \right\} \right) \\ &= (f - g) \left(\bigcup_{i=1}^k B_{z^{(i)}} \cup B_{(z_1^{(k+1)}, z_2^{(k)})} \right) \\ &= (f - g) \left(\bigcup_{i=1}^{k-1} B_{z^{(i)}} \cup B_{(z_1^{(k+1)}, z_2^{(k)})} \right) \end{aligned}$$

where the last equality follows from $(z_1^{(k)}, z_2^{(k)})$ being component-wise greater than or equal to $(z_1^{(k+1)}, z_2^{(k)})$. The inductive hypothesis implies that the quantity in the last line of the above derivation is ≥ 0 .

Case 2: $\zeta_{z_1^{(k+1)}}^{z_1^{(k)}}(w_2) \geq 0$ for all $w_2 \geq z_2^{(k+1)}$.

In this case, we have

$$\int_{z_2^{(k+1)}}^{q_2} \int_{z_1^{(k+1)}}^{z_1^{(k)}} (f - g) dz_1 dz_2 = \int_{z_2^{(k+1)}}^{q_2} \zeta_{z_1^{(k+1)}}^{z_1^{(k)}}(w) dw \geq 0.$$

Therefore, it follows that

$$\begin{aligned} (f - g)(A) &= (f - g) \left(\bigcup_{i=1}^k B_{z^{(i)}} \right) \\ &\quad + (f - g) \left(\left\{ z : z_1^{(k+1)} \leq z_1 \leq z_1^{(k)} \text{ and } z_2^{(k+1)} \leq z_2 \right\} \right) \\ &\geq (f - g) \left(\bigcup_{i=1}^k B_{z^{(i)}} \right) \geq 0, \end{aligned}$$

where the final inequality follows from the inductive hypothesis. □

6.3 Application: An Optimal Mechanism with an Infinite Menu Size

In this section, we will use Theorem 7 to calculate the optimal mechanism for two items distributed according to Beta distributions. In doing so we illustrate a general approach for finding closed-form descriptions of optimal mechanisms via the following steps: (i) definition of the sets S_{top} and S_{right} , (ii) computation of a critical price p^* , (iii) definition of a canonical partition in terms of (i) and (ii), and (iv) application of Theorem 7. Our approach succeeds in pinning down optimal mechanisms in all examples considered in Sections 6.3–6.4, and we expect it to be broadly applicable. Finally, it is noteworthy that the optimal mechanism for the setting studied in this section offers the buyer a menu of uncountably infinitely many lotteries to choose from. Using our approach we can nevertheless compute and succinctly describe the optimal mechanism. We also note in Remark 4 that our identified mechanism is essentially unique, hence the uncountability of the menu is inevitable.

Consider two items whose values are distributed independently according to the distributions $\text{Beta}(a_1, b_1)$ and $\text{Beta}(a_2, b_2)$, respectively. That is, the distributions are given by to the following two density functions on $[0, 1]$:

$$f_1(x) = \frac{1}{B(a_1, b_1)} x^{a_1-1} (1-x)^{b_1-1}; \quad f_2(y) = \frac{1}{B(a_2, b_2)} y^{a_2-1} (1-y)^{b_2-1}.$$

To find the optimal mechanism for our example setting, we first compute the measure μ induced by f . Notice that

$$\begin{aligned} -\nabla f(x, y) \cdot (x, y) - 3f(x, y) &= -x f_2(y) \frac{\partial f_1(x)}{\partial x} - y f_1(x) \frac{\partial f_2(y)}{\partial y} - 3f_1(x) f_2(y) \\ &= -(a_1 - 1) f_1(x) f_2(y) + (b_1 - 1) \frac{x}{1-x} f_1(x) f_2(y) \\ &\quad - (a_2 - 1) f_1(x) f_2(y) + (b_2 - 1) \frac{y}{1-y} f_1(x) f_2(y) - 3f_1(x) f_2(y) \\ &= f_1(x) f_2(y) \left(\frac{b_1 - 1}{1-x} + \frac{b_2 - 1}{1-y} + (1 - a_1 - b_1 - a_2 - b_2) \right) \end{aligned}$$

where the last equality used the identity $\frac{x}{1-x} = \frac{1}{1-x} - 1$. We also observe that $f_1(x)x = 0$ whenever $x = 0$ or $x = 1$ (as long as $b_1 > 1$), and an analogous property holds for y . Thus, the transformed measure μ is comprised of:

- a point mass of +1 at the origin; and
- mass distributed on $[0, 1]^2$ according to the density function

$$f_1(x) f_2(y) \left(\frac{b_1 - 1}{1-x} + \frac{b_2 - 1}{1-y} + (1 - a_1 - b_1 - a_2 - b_2) \right).$$

Note that in the case $b_i = 1$, our analysis still holds, except there is also positive mass on the boundary $x_i = 1$.

Deriving the Optimal Mechanism for a Concrete Setting of Parameters.

We now analyze a concrete example of two independent Beta distributed items where $a_1 = a_2 = 1$ and $b_1 = b_2 = 2$. That is, we consider two items whose values are

distributed independently according to the following two density functions on $[0, 1]$:

$$f_1(x) = 2(1 - x); \quad f_2(y) = 2(1 - y).$$

As discussed above, the transformed measure μ comprises:

- a point mass of +1 at the origin; and
- mass distributed on $[0, 1]^2$ according to the density function

$$f_1(x)f_2(y) \left(\frac{1}{1-x} + \frac{1}{1-y} - 5 \right).$$

Note that the density of μ is positive on $\mathcal{P} = \left\{ (x, y) \in (0, 1)^2 : \frac{1}{1-x} + \frac{1}{1-y} > 5 \right\} \cup \{\vec{0}\}$ and non-positive on $\mathcal{N} = \left\{ (x, y) \in [0, 1]^2 \setminus \{\vec{0}\} : \frac{1}{1-x} + \frac{1}{1-y} \leq 5 \right\}$, and that $\mathcal{N} \cup \{\vec{0}\}$ is a decreasing set.

Step (i). We first attempt to identify candidate functions for s_1 and s_2 that will lead to a well-formed canonical partition. We do this by defining two sets $S_{\text{top}}, S_{\text{right}} \subset [0, 1]^2$. We require that $(x, y) \in S_{\text{top}}$ iff $\int_y^1 \mu(x, t) dt = 0$. That is, starting from any point $z \in S_{\text{top}}$ and integrating the density of μ “upwards” from $t = y$ to $t = 1$ yields zero. Since $\mathcal{N} \cup \{\vec{0}\}$ is a decreasing set, it follows that $S_{\text{top}} \subset \mathcal{N}$ and that integrating μ upwards starting from any point above S_{top} yields a positive integral. Similarly, we say that $(x, y) \in S_{\text{right}}$ iff $\int_x^1 \mu(t, y) dt = 0$, noting that $S_{\text{right}} \subset \mathcal{N}$. S_{top} and S_{right} are shown in Figure 6-3.

We analytically compute that $(x, y) \in S_{\text{top}}$ if and only if $y = \frac{2-3x}{4-5x}$. Similarly, $(x, y) \in S_{\text{right}}$ if and only if $x = \frac{2-3y}{4-5y}$.

In particular, for any $x \leq 2/3$ there exists a y such that $(x, y) \in S_{\text{top}}$, and there does not exist such a y if $x > 2/3$. Furthermore, it is easy to verify by computing the second derivative of $\frac{\partial^2}{\partial x^2} \frac{2-3x}{4-5x} = -\frac{20}{(4-5x)^3} < 0$ that the region below S_{top} and the region below S_{right} are strictly convex.

Step (ii). We now need to calculate the critical point and the critical price. To do this we set the critical price $p^* \approx .5535$ as the intercept of the 45° line in Figure 6-3

which causes $\mu(Z) = 0$ for the set $Z \subset [0, 1]^2$ lying below S_{top} , S_{right} and the 45° line. We can also compute the critical point $(x_{\text{crit}}, y_{\text{crit}}) \approx (.0618, .0618)$ by finding the intersection of the critical price line with the sets S_{top} and S_{bottom} . Moreover, by the definition of the sets S_{top} and S_{bottom} , we know that the candidate boundary functions are $s_1(x) = \frac{2-3x}{4-5x}$ and $s_2(y) = \frac{2-3y}{4-5y}$, with domain $[0, x_{\text{crit}})$ and $[0, y_{\text{crit}})$ respectively.

Step (iii). We can now compute the canonical partition and decompose $[0, 1]^2$ into the following regions:

$$\mathcal{A} = \{(x, y) : x \in [0, x_{\text{crit}}) \text{ and } y \in [s_1(x), 1]\}; \mathcal{B} = \{(x, y) : y \in [0, y_{\text{crit}}) \text{ and } x \in [s_2(y), 1]\}$$

$$\mathcal{W} = \{(x, y) \in [x_{\text{crit}}, 1] \times [y_{\text{crit}}, 1] : x + y \geq p^*\}; Z = [0, 1]^2 \setminus (\mathcal{W} \cup \mathcal{A} \cup \mathcal{B})$$

as illustrated in Figure 6-3.

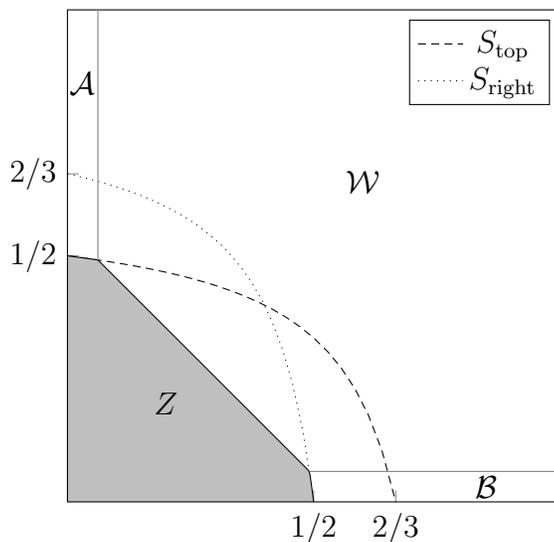


Figure 6-3: The well-formed canonical partition for $f_1(x) = 2(1 - x)$ and $f_2(y) = 2(1 - y)$.

Step (iv). We claim that the canonical partition $Z \cup \mathcal{A} \cup \mathcal{B} \cup \mathcal{W}$ is well-formed with respect to μ . Condition 2 is satisfied by construction of S_{top} and S_{right} and the corresponding discussion in Step (i). To check for Condition 1, note that given the definition of p^* , it holds that for all regions $R = Z, \mathcal{A}, \mathcal{B}$ and \mathcal{W} , we have $\mu(R) = 0$. Recall that $S_{\text{top}}, S_{\text{right}} \subset \mathcal{N}$ and, since $\mathcal{N} \cup \{\vec{0}\}$ is a decreasing set, μ has negative

density along these curves and all points below either curve, other than at the origin. Hence, $\mu_-|_Z \succeq_1 \mu_+|_Z$ which implies that $\mu|_Z \preceq_{cvx} 0$. Hence, the only non-trivial condition of Definition 15 that we need to verify is $\mu|_{\mathcal{W}} \succeq_2 0$. In fact, we can apply Lemma 11 to conclude the stronger dominance relation $\mu|_{\mathcal{W}} \succeq_1 0$.

We now how Lemma 11 can be applied for verifying that $\mu_+|_{\mathcal{W}} \succeq_1 \mu_-|_{\mathcal{W}}$. We set $\mathcal{C} = [x_{\text{crit}}, 1] \times [y_{\text{crit}}, 1]$ and $\mathcal{R} = Z \cap \mathcal{C}$, so that $\mathcal{W} = \mathcal{C} \setminus \mathcal{R}$. We let g and h being the positive and negative parts of the density function of $\mu|_{\mathcal{W}}$, respectively, so that the density of $\mu|_{\mathcal{W}}$ is given by $g - h$. Since Z lies below *both* curves S_{top} and S_{right} , we know that integrating the density of μ along any horizontal or vertical line outwards starting anywhere on the boundary of Z yields a non-positive quantity, verifying the second condition of Lemma 11. In addition, on $\mathcal{W} = \mathcal{C} \setminus \mathcal{R}$, we have

$$g(x, y) - h(x, y) = f_1(x)f_2(y) \left(\frac{1}{1-x} + \frac{1}{1-y} - 5 \right)$$

which satisfies the third condition of Lemma 11, as $1/(1-x) + 1/(1-y) - 5$ is increasing. Finally, we verify the first condition of Lemma 11 by integrating $g - h$ over \mathcal{C} . This integral is equal to $\mu(\mathcal{W}) = 0$ and thus all conditions of Lemma 11 are satisfied.

Having verified all conditions of Definition 15 we apply Theorem 7 to conclude that the optimal mechanism for this example is the following.

Example 3. *The optimal mechanism for selling two independent items whose values are distributed according to $f_1(x) = 2(1-x)$ and $f_2(y) = 2(1-y)$ has the following outcome for a buyer of type (x, y) :*

- If $(x, y) \in Z$, the buyer receives no goods and is charged 0.
- If $(x, y) \in \mathcal{A}$, the buyer receives item 1 with probability $-s'_1(x) = \frac{2}{(4-5x)^2}$, item 2 with probability 1, and is charged $s_1(x) - xs'_1(x) = \frac{2-3x}{4-5x} + \frac{2x}{(4-5x)^2}$.
- If $(x, y) \in \mathcal{B}$, the buyer receives item 2 with probability $-s'_2(y) = \frac{2}{(4-5y)^2}$, item 1 with probability 1, and is charged $s_2(y) - ys'_2(y) = \frac{2-3y}{4-5y} + \frac{2y}{(4-5y)^2}$.
- If $(x, y) \in \mathcal{W}$, the buyer receives both items and is charged $p^* \approx .5535$.

Remark 4. Note that the mechanism identified in Example 3 offers an uncountably large menu of lotteries. One could wonder whether there exists a different optimal mechanism offering a finite menu. Using our duality theorem we can easily argue that the utility function induced by every optimal mechanism equals the utility function $u(x)$ induced by our mechanism in Example 3. Hence, up to the choice of subgradients at the measure-zero set of types where $\nabla u(x)$ is discontinuous, the allocations offered by any optimal mechanism must agree with those of our mechanism in Example 3. Therefore, every optimal mechanism must offer an uncountably large menu. We show uniqueness of the optimal mechanism in Section 6.3.1

Summary of Beta Distributions. Example 3 shows that the optimal mechanism for two Beta distributed items offers a continuum of lotteries, thereby having infinite menu-size complexity [HN13]. Still, using our techniques we can obtain a succinct and easily-computable description of the mechanism.

Working similarly to Example 3, we can obtain the optimal mechanism for broader settings of parameters. Figure 6-4 illustrates the optimal mechanism for two items distributed according to Beta distributions with different parameters. The reader can experiment with different settings of parameters at [Tza].



Figure 6-4: Canonical Partitions for different cases of Beta distributions. The shaded region is where the measure μ becomes negative. (Note that when the second parameter b_i of the Beta distribution of some item i equals 1, μ has positive mass on the outer boundary $x_i = 1$.) (1) Beta(1,1) and Beta(1,1), (2) Beta(2,2) and Beta(1,1), (3) Beta(2,2) and Beta(2,2).

6.3.1 Uniqueness of Mechanism in Example 3

To argue that the utility $u(x)$ is shared by all optimal mechanisms, we start by constructing an optimal solution γ^* to the RHS of (4.1). γ^* needs to satisfy the complementary slackness conditions of Corollary 1 against any optimal solution u^* to the LHS of (4.1). We will choose our solution γ^* so that the complementary slackness conditions will imply $u^* = u$. Let us proceed with the choice of γ^* . Recall the canonical partition $Z \cup \mathcal{A} \cup \mathcal{B} \cup \mathcal{W}$ of the type space, identified above, and illustrated in Figure 6-3. We define a solution γ^* to the RHS of (4.1) that separates into the four regions as follows (the optimality of this γ^* follows easily by checking that it satisfies the complementary slackness conditions of Corollary 1 against u):

Region Z Recall that, in region Z , we have $\mu|_Z \preceq_{cvx} 0$. Our solution γ^* matches the +1 unit of mass sitting at the origin to the negative mass spread throughout region Z , by moving positive mass to coordinate-wise larger points and performing mean preserving spreads. By the complementary slackness conditions of Corollary 1 (see Remark 1 for intuition), it follows that $u^*(x) = 0$, for any optimal solution u^* to the LHS of (4.1).

Regions \mathcal{A} and \mathcal{B} In regions \mathcal{A} and \mathcal{B} our solution γ^* transports mass vertically and, respectively, horizontally. The complementary slackness conditions imply then that any optimal solution u^* to the LHS of (4.1) u^* must change linearly in the second coordinate in region \mathcal{A} and linearly in the first coordinate in region \mathcal{B} .

Region \mathcal{W} Finally, in region \mathcal{W} we want to show that any optimal u satisfies $|u(\vec{x}) - u(\vec{y})| = \|\vec{x} - \vec{y}\|_1$ if $\vec{x} \geq \vec{y}$ coordinate-wise. This is not as straightforward as the previous 2 cases as we don't have an explicit description of the optimal dual solution. However, we can use Lemma 11 to show that there exists a measure γ^* which is optimal for the dual and matches types on the top right corner (with values $\approx (1, 1)$) to types close to the bundling line (with values $x_1 + x_2 \approx p^*$) which implies that any optimal function u must be linear in \mathcal{W} .

By continuity, any optimal u must be equal to $z_1 + z_2 - p^* = 0$ when $z_1 + z_2 = p^*$. Moreover, it holds that $u(z) \leq z_1 + z_2 - p^*$, because u is 1-Lipschitz. We will now show the reverse inequality by showing that $u(1, 1) = 2 - p^*$. Recall that the density of measure μ in region \mathcal{W} is equal to:

$$\mu(z_1, z_2) = f_1(z_1)f_2(z_2) \left(\frac{1}{1-z_1} + \frac{1}{1-z_2} - 5 \right)$$

where $f_1(x) = f_2(x) = (1-x)$. Lemma 11 implied that $\mu_+|_{\mathcal{W}} \succeq_1 \mu_-|_{\mathcal{W}}$ but didn't give a transport map γ constructively. To partially specify a transport map γ that is optimal for the dual, we define for sufficiently small $\epsilon > 0$ the measure μ' which has density

$$\mu'(z_1, z_2) = f_1(z_1)f_2(z_2) \left(\frac{1}{\epsilon} + \max \left(\frac{1}{1-z_2}, \frac{1}{1-z_1} \right) - 5 \right)$$

when $(z_1, z_2) \in [1-\epsilon, 1]^2$ and $\mu'(z_1, z_2) = \mu(z_1, z_2)$ otherwise. In particular, μ' is obtained by removing some positive mass from μ in $[1-\epsilon, 1]^2$ and thus $\mu'(\mathcal{W}) < \mu(\mathcal{W}) = 0$. Moreover, notice that we defined μ' so that $\frac{\mu'(z_1, z_2)}{f_1(z_1)f_2(z_2)}$ is still an increasing function. Now, let R' be the region enclosed within the curves $s_1(x)$, $s_2(y)$, $x+y = p^*$ and $x+y = p'$ for $p' > p^*$ so that $\mu'(\mathcal{W} \setminus R') = 0$. This defines a decomposition of measure $\mu|_{\mathcal{W}}$ into two measures $\mu'|_{\mathcal{W} \setminus R'}$ and $\mu|_{\mathcal{W}} - \mu'|_{\mathcal{W} \setminus R'}$ of zero total mass (Figure 6-5).

We apply Lemma 11 for μ' in region $\mathcal{W} \setminus R'$ to get that $\mu'|_{\mathcal{W} \setminus R'} \succeq_1 0$. We also have that $(\mu - \mu')|_{\mathcal{W}} \succeq_1 \mu|_{R'}$ since $(\mu - \mu')|_{\mathcal{W}}$ contains only positive mass supported on $[1-\epsilon, 1]^2$ and every point in the support pointwise dominates every point in the support of $\mu|_{R'}$. Thus, there exists an optimal transport map γ^* in region \mathcal{W} such that $\gamma^* = \gamma^{(i)} + \gamma^{(ii)}$ and $\gamma^{(i)}$ transports the mass $\mu'|_{\mathcal{W} \setminus R'}$ while $\gamma^{(ii)}$ transports mass arbitrarily from $(\mu - \mu')|_{\mathcal{W}}$ to $\mu|_{R'}$. Given such an optimal γ^* , the complementary slackness conditions of Corollary 1 imply that any feasible u must satisfy $|u(\vec{z}) - u(\vec{z}')| = \|\vec{z} - \vec{z}'\|_1$ whenever mass is transferred from \vec{z} to \vec{z}' . This can only happen if $u(1, 1) = 2 - p^*$ and implies that $u(\vec{z}) = z_1 + z_2 - p^*$ everywhere on \mathcal{W} .

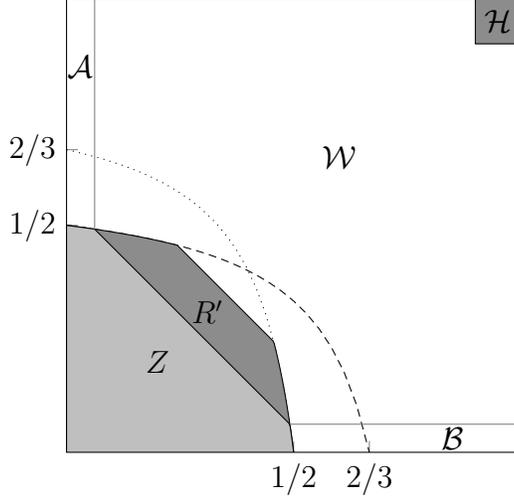


Figure 6-5: Decomposition of measure $\mu|_{\mathcal{W}}$ into measures $\mu'|_{\mathcal{W}\setminus R'}$ and $\mu|_{\mathcal{W}} - \mu'|_{\mathcal{W}\setminus R'}$. The dark shaded regions R' and $\mathcal{H} = [1 - \epsilon, 1]^2$ show the support of $\mu|_{\mathcal{W}} - \mu'|_{\mathcal{W}\setminus R'}$.

6.4 More Applications: Distributions of Unbounded Support

So far, our results have focused on type distributions with bounded support. Even though we do not know extensions of our strong duality theorem (Theorem 2), and the optimal menu conditions (Theorem 3) for unbounded type distributions, due to technical issues, the rest of our results can easily be generalized. In this section, we briefly discuss how Theorem 1, Lemma 1, and Theorem 7 can be modified to accommodate settings with unbounded type spaces, as long as the type distribution decays sufficiently rapidly towards infinity.

We can often obtain a “transformed measure” (analogous to Theorem 1 even when type spaces are unbounded) using the divergence theorem or simply integration by parts. We need to ensure, however, that the density function f decays sufficiently quickly so that there is no “surface term at infinity.” For example, we may require that $\lim_{z_i \rightarrow \infty} f_i(z_i)z_i^2 \rightarrow 0$, as in [DDT13]. We note that without some conditions on the decay rate of f , it is possible that the supremum revenue achievable is infinite and thus no optimal mechanism exists.

Similar issues arise when integrating with respect to an unbounded measure μ .

It is helpful therefore to consider only measures μ such that $\int \|x\|_1 d|\mu| < \infty$, to ensure that $\int u d\mu$ is finite for any utility function u . The measures in our examples satisfy this property. We can (informally speaking) attempt to extend this definition to unbounded measures (with regularity conditions such as $\int \|x\|_1 d|\mu| < \infty$) by ensuring that whenever the “smaller” side has infinite value, so does the larger side.

Importantly, the calculations of Lemma 1 (weak duality) hold for unbounded μ , provided $\int \|x\|_1 d|\mu| < \infty$. Thus, tight certificates still certify optimality, even in the unbounded case. However, our strong duality proof relies on technical tools which require compact spaces, and thus these proofs do not immediately apply when μ is unbounded.

To summarize our discussion so far, we can often transform measures and obtain an analogue of Theorem 1 for unbounded distributions (provided the distributions decay sufficiently quickly), and can easily obtain a weak duality result for such unbounded measures, but additional work is required to prove whether strong duality holds.

6.4.1 An Example with Power-Law Distributions

In Example 4, the optimal mechanism for selling two power-law items is a grand bundling mechanism. The canonical partition induced by the exclusion set of the grand-bundling mechanism is degenerate (regions \mathcal{A} and \mathcal{B} are empty), and establishing the optimality of the mechanism amounts to establishing that the first-order stochastic dominance condition for the induced measure μ holds in region \mathcal{W} .

Example 4. *The optimal IC and IR mechanism for selling two items whose values are distributed independently according to the probability densities $f_1(x) = 5/(1+x)^6$ and $f_2(y) = 6/(1+y)^7$ respectively is a take-it-or-leave-it offer of the bundle of the two goods for price $p^* \approx .35725$.*

As discussed above, we can apply integration by parts or the divergence theorem as in Theorem 1 to obtain the transformed measure μ . It comprises:

- a point mass of +1 at the origin; and

- mass distributed on $[0, \infty)^2$ according to the density function

$$f_1(x)f_2(y) \left(\frac{6x}{1+x} + \frac{7y}{1+y} - 3 \right).$$

To prove that the proposed grand bundling mechanism is optimal, it suffices to show that $\mu_+|_{\mathcal{W}} \succeq_1 \mu_-|_{\mathcal{W}}$ and that $\mu_+|_Z \preceq_1 \mu_-|_Z$, where $\mathcal{W} = \{(x, y) | x + y \geq p^*\}$ and $Z = \{(x, y) | x + y < p^*\}$. The latter is trivially true as the only positive mass in region Z is at the origin. We will use Theorem 11 to prove the former. We apply the theorem for a function g corresponding to the density of $\mu_+|_{\mathcal{W}}$ and a function h corresponding to the density of $\mu_-|_{\mathcal{W}}$.

As p^* was computed so that $\mu_+(\mathcal{W}) = \mu_-(\mathcal{W})$, it is easy to see that the first two conditions are satisfied.

The fourth condition of Theorem 11 is satisfied with $\alpha(x) = f_1(x)$, $\beta(y) = f_2(y)$, and $\eta(x, y) = \frac{6x}{1+x} + \frac{7y}{1+y} - 3$, noting that η is indeed an increasing function.

All that remains is to verify the third condition of Theorem 11. We break this verification into two parts, depending on whether we are integrating with respect to x or y .

- We begin by considering integration with respect to x . That is, for any fixed $0 \leq y \leq p^*$, we must prove that

$$\int_{p^*-y}^{\infty} \left(\frac{6x}{1+x} + \frac{7y}{1+y} - 3 \right) \frac{5}{(1+x)^6} \cdot \frac{6}{(1+y)^7} dx \leq 0.$$

This integral evaluates to

$$\frac{-0.18565 + 1.1145y - 2y^2}{(1.35725 - y)^6(1+y)} \cdot \frac{6}{(1+y)^7}.$$

Since the second term is positive, it suffices to prove that the first term is negative. The denominator of the first term is always positive so it suffices to prove that its numerator is negative. Indeed, the numerator is maximized at $y = .2786$, in which case the numerator evaluates to $-.0304$.

- We now consider integration with respect to y . Analogously to the computation above, for any fixed $0 \leq x \leq p^*$ we must prove that

$$\int_{p^*-x}^{\infty} \left(\frac{6x}{1+x} + \frac{7y}{1+y} - 3 \right) \cdot \frac{6}{(1+y)^7} dx \leq 0.$$

This integral evaluates to

$$\frac{-0.0951667 + .595416x - 1.66667x^2}{(1.35725 - x)^7(1+x)}.$$

As before, it suffices to prove that the numerator is negative. We verify that, indeed, the numerator achieves its maximum at $x = .178625$, in which case the numerator is $-.0419886$.

Therefore, we have proven, by Theorem 11, that $\mu_{+|\mathcal{W}} \succeq_1 \mu_{-|\mathcal{W}}$, as desired.

6.4.2 An Example with Exponential Distributions

Our final example, Example 5, provides a complete solution for the optimal mechanism for two items distributed according to independent exponential distributions. In this case, the canonical partition induced by the exclusion set of the mechanism is missing region \mathcal{A} , and possibly region \mathcal{B} (if $\lambda_1 = \lambda_2$).

Example 5. *For all $\lambda_1 \geq \lambda_2 > 0$, the optimal IC and IR mechanism for selling two items whose values are distributed independently according to exponential distributions f_1 and f_2 with respective parameters λ_1 and λ_2 offers the following menu:*

1. *receive nothing, and pay 0;*
2. *receive the first item with probability 1 and the second item with probability λ_2/λ_1 , and pay $2/\lambda_1$; and*
3. *receive both items, and pay p^* ;*

where p^* is the unique $0 < p^* \leq 2/\lambda_2$ such that

$$\mu(\{(x, y) \in \mathbb{R}_{\geq 0}^2 : x + y \leq p^* \text{ and } \lambda_1 x + \lambda_2 y \leq 2\}) = 0,$$

where μ is the transformed measure of the joint distribution.

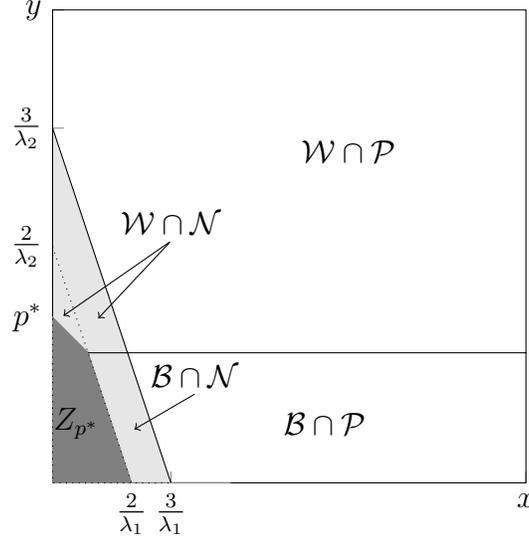


Figure 6-6: The canonical partition of $\mathbb{R}_{\geq 0}^2$ for the proof of Example 5. In this diagram, $p^* > 2/\lambda_1$. If $p^* \leq 2/\lambda_1$, \mathcal{B} is empty. The positive part μ_+ of μ is supported inside $\mathcal{P} \cap \{\vec{0}\}$ while the negative part μ_- is supported within $Z_{p^*} \cup \mathcal{N}$.

The transformed measure μ comprises:

- a point mass of +1 at the origin; and
- mass distributed on $[0, \infty)^2$ according to the density function

$$f_1(x)f_2(y) (\lambda_1 x + \lambda_2 y - 3).$$

A useful feature of independent exponential distributions is that integrating μ outwards along any line starting from (x, y) such that $\lambda_1 x + \lambda_2 y = 2$ yields 0.

Claim 8. For any vector $\vec{v} \in \mathbb{R}_{\geq 0}^2$, the sign of

$$\int_0^\infty (\lambda_1(x + \tau v_1) + \lambda_2(y + \tau v_2) - 3) f_1(x + \tau v_1) f_2(y + \tau v_2) d\tau,$$

is equal to the sign of $\lambda_1 x + \lambda_2 y - 2$.

Proof. The claim follows by noting that:

$$\begin{aligned}
& \int_0^\infty (\lambda_1(x + \tau v_1) + \lambda_2(y + \tau v_2) - 3) f_1(x + \tau v_1) f_2(y + \tau v_2) d\tau \\
&= - \left. \frac{e^{-\lambda_1(x + \tau v_1) - \lambda_2(y + \tau v_2)} (\lambda_1(x + \tau v_1) + \lambda_2(y + \tau v_2) - 2)}{\lambda_1(x + \tau v_1) + \lambda_2(y + \tau v_2)} \right|_{\tau=0}^\infty \\
&= \frac{e^{-\lambda_1 x - \lambda_2 y} (\lambda_1 x + \lambda_2 y - 2)}{\lambda_1 x + \lambda_2 y}.
\end{aligned}$$

□

We now show that Claim 8 can be used to prove that the canonical partition shown in Figure 6-6 is well-formed. In region \mathcal{B} , Claim 8 implies that $\mu|_{\mathcal{B}}([x, \infty) \times [y, y + \epsilon]) \geq 0$ for all $\lambda_1 x + \lambda_2 y \geq 2$ with equality if $\lambda_1 x + \lambda_2 y = 2$.

To complete the proof, it suffices to show that $\mu_+|_{\mathcal{W}} \succeq_1 \mu_-|_{\mathcal{W}}$ and that $\mu_+|_Z \preceq_1 \mu_-|_Z$, where $\mathcal{W} = \{(x, y) | x + y \geq p^*\}$ and $Z = \{(x, y) | x + y < p^*\}$. The latter is trivially true as the only positive mass in region Z is at the origin. We will use Theorem 11 to prove the former. We apply the theorem for a function g corresponding to the density of $\mu_+|_{\mathcal{W}}$ and a function h corresponding to the density of $\mu_-|_{\mathcal{W}}$.

As p^* was computed so that $\mu_+(\mathcal{W}) = \mu_-(\mathcal{W})$, it is easy to see that the first two conditions are satisfied. The third condition follows directly from Claim 8, while the fourth condition of Theorem 11 is satisfied with $\alpha(x) = f_1(x)$, $\beta(y) = f_2(y)$, and $\eta(x, y) = \lambda_1 x + \lambda_2 y - 3$, noting that η is indeed an increasing function.

Chapter 7

Complexity of Mechanism Design

While our duality approach presented so far ensures that every optimal mechanism has a dual certificate proving its optimality, it does not address the computational task of actually finding the optimal mechanism. As mentioned in Chapter 2, there are several algorithmic approaches that compute optimal mechanisms for one or more bidders [CDW12a, CDW12b, AFH⁺12]. Those results apply when bidders' valuations are given explicitly, by listing every valuation in their support together with the probability it appears. Such a manner of describing a joint distribution, however, is often unnatural. For example, in the case of independently distributed items, explicitly listing the joint distribution can be exponentially longer than listing each marginal distribution separately.

Our goal in this chapter is to investigate the complexity of computing the optimal mechanism in cases where the distribution of values can be represented more succinctly. We find that even in very simple settings with a single additive bidder with independent valuations, the problem becomes computationally intractable. To prove this, we focus on instances with discrete distributions and characterize the optimal mechanism for a sufficiently large class of input instances. We then use those instances to reduce a computationally hard problem to the problem of optimal mechanism design. Our main result is the following:

Theorem 8. *There is no expected polynomial-time solution to the optimal mechanism*

design problem (formal definition in Section 7.1) unless $ZPP \supseteq P^{\#P}$.

In particular, it is $\#P$ -hard to determine whether every optimal mechanism assigns a specific item to a specific type of bidder with probability 0 or with probability 1, given the promise that one of these two cases holds simultaneously for all optimal mechanisms.

The above are true even in the case of a single additive, quasi-linear bidder, whose values for the items are independently distributed on two rational numbers with rational probabilities.

This result states that even in a simple class of instances, computing the optimal mechanism is computationally intractable. This suggests that optimal mechanisms can have a very rich and combinatorial structure.

Before we proceed to present the proof of Theorem 8, we revisit the formulation of the optimal mechanism design problem more formally to obtain a lower bound that is as broad as possible.

7.1 The Optimal Mechanism Design Problem

As we are aiming for a broad lower bound we revisit the definition of the *Optimal Mechanism Design (OMD) problem* placing no constraints on the form of the sought after mechanism, or how this mechanism is meant to be described. Hence, our lower bounds apply to computing direct revelation mechanisms as well as any conceivable type of mechanism. Moreover, we are able to identify a simple enough class of instances consisting of a single bidder with additive valuations for every item that are drawn independently from distributions of support 2. Our lower-bound directly applies to cases with many bidders and more complicated valuation functions other than additive.

We now describe broadly the optimal mechanism design problem.

INPUT: This consists of the names of the items and the bidders, the allocation constraints of the setting (specifying, e.g., that an item may be allocated to at most one bidder, etc.), and a probability distribution on the valuations, or *types*, of the

bidders. The type of a bidder incorporates information about how much he values every subset of the items, as well as what utility he derives for receiving a subset at a particular price. For example, the type of an additive quasi-linear bidder can be encapsulated in a vector of values (one value per item). We won't make any assumptions about how the allocation constraints are specified. In general, these could either be hard-wired to a family of instances of the OMD problem, or provided as part of the input in a computationally meaningful way. For the purposes of our intractability results, the allocation constraints will be trivial, enforcing that we can only allocate at most one copy of each item, and we restrict our attention to instances with precisely these allocation constraints. As far as the type distribution is concerned, we restrict our attention to additive quasi-linear bidders with independent values for the items. So, for our purposes, the type distribution of a bidder is specified by specifying its marginal on each item. We assume that each marginal is given explicitly, as a list of the possible values for the item as well as the probabilities assigned to each value.¹

DESIRED OUTPUT: The goal is to compute a (possibly randomized) auction that optimizes, over all possible auctions, the expected revenue of the auctioneer, i.e. the expected sum of prices paid by the bidders at the Bayes Nash equilibrium of the auction,² where the expectation is taken with respect to the bidders' types, the randomness in their strategies (if any) at the Bayes Nash equilibrium, as well as any internal randomness that the auction uses.

We note that there is a large universe of possible auctions with widely varying formats, e.g. sealed envelope auctions, dynamic auctions, all-pay auctions, etc. And

¹There are of course other ways to describe these marginals. For example, we may only have sample access to them, or we may be given a circuit that takes as input a value and outputs the probability assigned to that value. As our goal is to prove lower bounds, the assumption that the marginals are provided explicitly in the input only makes the lower bounds stronger.

²Informally *Bayesian Nash equilibrium* is the extension of *Nash equilibrium* to incomplete-information games, i.e. games where the utilities of players are sampled from a probability distribution. We won't provide a formal definition as it is quite involved and is actually not required for our lower bounds, which focus on the single-bidder case.

For the purposes of the problem definition though, we note that, if an auction has multiple Bayesian Nash equilibria, its revenue is not well-defined as it may depend on what Bayesian Nash equilibrium the bidders end up playing. So we would like to avoid such auctions given the uncertainty about their revenue. Again this complication won't be relevant for our results as all auctions we construct in our hardness proofs will have a unique Bayes Nash equilibrium as they are for single bidder instances.

there could be different auctions with optimal expected revenue. As our goal is to prove robust intractability results for OMD, we take a general approach imposing no restrictions on the format of the auction, and no restrictions on the way the auction is encoded. The encoding should specify in a computationally meaningful way what actions are available to the bidders, how the items are allocated depending on the actions taken by the bidders, and what prices are charged to them, where both allocation and prices could be outputs of a randomized function of the bidders' actions. In particular, a computationally efficient solution to OMD induces the following:

Auction Computation & Simulation: A computationally efficient solution to a family \mathcal{I} of OMD problems induces a pair of algorithms \mathcal{C} and \mathcal{S} satisfying the following:

1. [*auction computation*] $\mathcal{C} : \mathcal{I} \rightarrow \mathcal{E}$ is an expected polynomial-time algorithm mapping instances $I \in \mathcal{I}$ of the OMD problem to auction encodings $\mathcal{C}(I) \in \mathcal{E}$; e.g. $\mathcal{C}(I)$ may be “second price auction”, or “English auction with reserve price \$5”, etc.
2. [*auction simulation*] \mathcal{S} is an expected polynomial-time algorithm mapping instances $I \in \mathcal{I}$ of the OMD problem, encodings $\mathcal{C}(I)$ of the optimal auction for I , and realized types t_1, \dots, t_m for the bidders, to a sample from the (possibly randomized) allocation and price rule of the auction encoded by $\mathcal{C}(I)$, at the Bayes Nash equilibrium of the auction when the types of the bidders are t_1, \dots, t_m .

Clearly, Property 1 holds because computing the optimal auction encoding $\mathcal{C}(I)$ for an instance I is assumed to be efficient. But why does there exist a simulator \mathcal{S} as in Property 2? Well, when the auction $\mathcal{C}(I)$ is executed, then somebody (either the auctioneer, or the bidders, or both) need to do some computation: the bidders need to decide how to play the auction (i.e. what actions from among the available ones to use), and then the auctioneer needs to allocate the items and charge the

bidders. In the special case of a direct Bayesian Incentive Compatible mechanism,³ the bidders need not do any computation, as it is a Bayes Nash equilibrium strategy for each of them to truthfully report their type to the auctioneer. In this case, all the computation must be done by the auctioneer who needs to sample from the (possibly randomized) allocation and price rule of the mechanism given the bidders’ reported types. In general (possibly non-direct, or multi-stage) mechanisms, both bidders and auctioneer may need to do some computation: the bidders need to compute their Bayes Nash equilibrium strategies given their types, and the auctioneer needs to sample from the (possibly random) allocation and price rule of the mechanism given the bidders’ strategies. These computations must all be computationally efficient, as otherwise the execution of the auction $\mathcal{C}(I)$ would not be computationally efficient. Hence an efficient solution must induce an efficient simulator \mathcal{S} .

Remark 5. We note that a proper study of the computational complexity of OMD cannot drop the requirement that auction simulation be efficient. Without this requirement, the auction computation which always outputs “the input itself is an encoding of the optimal mechanism for that input” would spuriously be considered an efficient solution for every OMD instance. Furthermore, notice that \mathcal{S} simulates the combined computations performed by both the bidders and the auctioneer in an execution of the auction. It is important that the bidders’ computations can be simulated efficiently, as placing no computational restrictions on the bidders leads to spurious “efficient” solutions to OMD, as discussed in Section 7.6.2.

In view of the above discussion, Theorem 8 establishes that, even in very simple special cases of the OMD problem, there does not exist a pair $(\mathcal{C}, \mathcal{S})$ of efficient auction computation and simulation algorithms, i.e. the optimal auction cannot be found computationally efficiently, or cannot be executed efficiently, or both. We note that our hardness result is subject to the assumption $\text{ZPP} \not\subseteq \text{P}^{\#\text{P}}$ (rather than $\text{P} \not\subseteq \text{P}^{\#\text{P}}$) solely because we prove lower bounds for randomized mechanisms.

³A mechanism is *direct* if the only available actions to a bidder is to report a type in the support of his type-distribution. A direct mechanism is *Bayesian Incentive Compatible* if it is a Bayes Nash equilibrium for every bidder to truthfully report his type to the auctioneer.

Remark 6 (Hardness of IC Mechanisms). A lot of research on optimal mechanism design has focused on finding optimal Bayesian Incentive Compatible (IC) mechanisms, as focusing on such mechanisms costs nothing in revenue due to the direct revelation principle (see [NRTV07] and Section 7.2). As an immediate corollary of Theorem 8 we obtain that it is #P-hard to compute the (possibly randomized) allocation and price rule of the optimal IC mechanism. However, Theorem 8 is much broader, in two respects: a. in the definition of the OMD problem we impose no constraints on what type of auction should be found; b. we don't require an explicit computation of the (possibly randomized) allocation and price rule of the mechanism, but allow an expected polynomial-time algorithm that samples from the allocation and price rule.

A mechanism \mathcal{M} for an instance I specifies a set \mathcal{A} of actions available to the bidder together with a rule mapping each action $\alpha \in \mathcal{A}$ to a (possibly randomized) allocation $A_\alpha \in \{0, 1\}^n$, specifying which items the bidder gets, and a (possibly randomized) price $\tau_\alpha \in \mathbb{R}$ that the bidder pays, where A_α and τ_α could be correlated. Facing this mechanism, a bidder whose values for the items are instantiated to some vector \vec{v} chooses any action in $\arg \max_{\alpha \in \mathcal{A}} \{\vec{v} \cdot \mathbb{E}(A_\alpha) - \mathbb{E}(\tau_\alpha)\}$ or any distribution on these actions, as long as the maximum is non-negative, since $\vec{v} \cdot \mathbb{E}(A_\alpha) - \mathbb{E}(\tau_\alpha)$ is the expected utility of the bidder for choosing action α .⁴ In particular, any such choice is a *Bayesian Nash equilibrium* behavior for the bidder, in the degenerate case of a single bidder that we consider. If for all vectors \vec{v} there is a unique optimal action $\alpha_{\vec{v}} \in \mathcal{A}$ in the above optimization problem, then mechanism \mathcal{M} induces a mapping from valuations \vec{v} to (possibly randomized) allocation and price pairs $(A_{\alpha_{\vec{v}}}, \tau_{\alpha_{\vec{v}}})$. If there are \vec{v} 's with non-unique maximizers, then we break ties in favor of the action with the highest $\mathbb{E}(\tau_\alpha)$ and, if there are still ties, lexicographically beyond that.⁵

⁴If the maximum utility under \vec{v} is negative, the bidder would “stay home.” To ease notation, we can include in \mathcal{A} a special action “stay home” that results in the bidder getting nothing and paying nothing. If all other actions give negative utility the bidder could just use this special action.

⁵We can enforce this tie-breaking with an arbitrarily small hit on revenue as follows: For all α , we decrease the (possibly random price) τ_α output by \mathcal{M} by a well-chosen amount—think of it as a rebate—which gets larger as $\mathbb{E}(\tau_\alpha)$ gets larger. We can choose these rebates to be sufficiently small so that they only serve the purpose of tie-breaking. These rebates won't affect our lower bounds.

In our hardness proof, we focused on direct mechanisms, in which each player’s action space coincides with his valuation space. However, it is straightforward to translate this hardness result to general mechanisms by making the following observation, called the “Revelation Principle” [NRTV07]:

- Any (possibly non-direct) mechanism \mathcal{M} has an equivalent IC, IR, direct mechanism \mathcal{M}' so that the two mechanisms induce the exact same mapping from types \vec{v} to (possibly randomized) allocation and price pairs. Indeed, given the (possibly randomized) allocation rule A and price rule τ of \mathcal{M} we can define the mapping $\vec{v} \mapsto (A_{\alpha_{\vec{v}}}, \tau_{\alpha_{\vec{v}}})$, where $\alpha_{\vec{v}}$ is an optimal action for a bidder of type \vec{v} . This mapping is in fact itself a IC, IR direct mechanism \mathcal{M}' . Clearly, \mathcal{M} and \mathcal{M}' have the same expected revenue.

As mentioned previously, the hard instances we narrow into in our proof satisfy that their optimal IC, IR direct mechanism is unique and it is a #P-hard problem to tell whether $p_{i^*}(S^*) = 0$ or 1 for a special item i^* and a special type S^* . The above observation implies that, for any instance in our family, any (possibly non-direct) optimal mechanism \mathcal{M} for this instance needs to induce an optimal direct mechanism. Since the latter is unique, \mathcal{M} needs to give the same (possibly randomized) allocation to type S^* that the unique direct mechanism does. Getting one sample from this allocation allows us to decide whether $p_{i^*}(S^*) = 1$, a #P-hard problem. As any efficient solution to our family of OMD instances would allow us to get samples from the allocation rule of an optimal mechanism in expected polynomial-time, we would be able to solve a #P-hard problem in expected polynomial-time. This establishes Theorem 8 (for unrestricted mechanisms).

Note that, for a meaningful lower bound, we cannot ask for $p(S), \tau(S)$ for all S , as there are too many S ’s—namely 2^n . Instead we need to ask for some implicit but computationally meaningful description of them, such as in the form of circuits, which can be evaluated on an input S in time polynomial in their size and the number of bits required to describe S —if we don’t require that p and τ can be evaluated efficiently on any given S we would allow for trivial solutions such as “ I is itself an implicit

description of the optimal p and τ for I .” We conclude with the following remark.

Remark 7. For the single-bidder instances we consider here, a direct mechanism that is Bayesian Incentive Compatible is also *Incentive Compatible* and vice versa. As all our hardness results are for single-bidder instances, they simultaneously show the intractability of computing optimal Bayesian Incentive Compatible as well as optimal Incentive Compatible mechanisms.

7.1.1 Our Approach

There are serious obstacles in establishing intractability results for optimal mechanism design, the main one being that the structure of optimal mechanisms is poorly understood even in very simple settings. To prove Theorem 8, we need to find a family of mechanism design instances whose optimal solutions are sufficiently complex to enable reductions from some #P-hard problem, while at the same time are sufficiently restricted so that solutions to the #P-hard problem can actually be extracted from the structure of the optimal mechanism.

However, pinning down such a family of sufficiently interesting instances of the problem for which we can still characterize the form of the optimal mechanism seems challenging as there is no apparent structure in the optimal mechanism even in the simple case of a single additive bidder. See the relevant discussion in Chapter 2 for several unintuitive examples. We follow a principled approach starting with a folklore, albeit exponentially large, linear program for revenue optimization, constructing a relaxation of this LP, and showing that, in a suitable class of instances, the solution of the relaxed LP is also a solution to the original LP, which has rich enough structure to embed a #P-hard problem. In more detail, our approach is the following:

- In Section 7.3.1 we present LP1, the folklore, albeit exponentially large, linear program for computing a revenue optimal auction.
- In Section 7.3.2 we relax the constraints of LP1 to construct a new, still exponentially large, linear program LP2. The solutions of the relaxed LP need not provide solutions to the original mechanism design problem. We prove however

that an optimal LP2 solution is indeed an optimal LP1 solution if it happens to be monotone and supermodular.

- In Section 7.3.3 we take LP3, the dual program to LP2. We interpret its solutions as solutions to a minimum-cost flow problem on a lattice.
- In Section 7.4.1 we characterize a *canonical solution* to a specific subclass of LP3 instances. This solution requires ordering of the subset sums of an appropriate set of integers.
- In Section 7.4.2 we use duality to convert a canonical LP3 solution to a unique LP2 solution. We are therefore able to characterize the unique solution for a variety of LP2 instances.
- In Section 7.4.3 we show that the LP2 solutions obtained above are also feasible and optimal for the corresponding LP1 instance. Thus, we gain the ability to characterize unique optimal solutions of a class of LP1 instances.
- In Section 7.5 we show how to encode a #P-hard problem into the class of LP1 instances that we have developed.

7.2 A Simple Class of Instances

Following Section 3.1, we consider settings with a single bidder with additive valuations. To establish a general lower bound we consider a simple instance with independent value distributions for every item i that are supported on a discrete set $\{a_i, a_i + d_i\}$. In particular, the bidder values item i at a_i , with probability $1 - q_i$, and at $a_i + d_i$, with probability q_i , independently of the other items, where a_i , d_i , and q_i are positive rational numbers. If he values i at a_i , we say that his value for i is “low” and, if he values it at $a_i + d_i$, we say that his value is “high.” The specification of the instance comprises the numbers $\{a_i, d_i, q_i\}_{i=1}^n$.

The lower-bound that we will show also applies to non-direct mechanisms that allow any interaction between the seller and the buyer. A mechanism \mathcal{M} in that case

specifies a set \mathcal{A} of actions available to the bidder together with a rule mapping each action $\alpha \in \mathcal{A}$ to a (possibly randomized) allocation $A_\alpha \in \{0, 1\}^n$, indicating which items the bidder gets, and a (possibly randomized) price $\tau_\alpha \in \mathbb{R}$ that the bidder pays, where A_α and τ_α could be correlated.

However, it is convenient to first study direct mechanisms as in Section 3.1, where the action set available to the bidder coincides with his type space $\times_i \{a_i, a_i + d_i\}$, i.e. the bidder declares his type to the mechanism. In this case, we can equivalently think of the actions available to the bidder as declaring any subset $S \subseteq N$, where the correspondence between subsets and value vectors is given by $\vec{v}(S) = \sum_{i \in N} a_i e_i + \sum_{i \in S} d_i e_i$ where e_i is the unit vector in dimension i ; i.e. S represents the items whose values are high. For all actions $S \subseteq N$, the mechanism induces a vector $\vec{p}(S) \in [0, 1]^n$ of probabilities that the bidder receives each item and an expected price $\tau(S) \in \mathbb{R}$ that the bidder must pay. The expected utility of a bidder of type $\vec{v}(S)$ for choosing action S' is given by $\vec{v}(S) \cdot \vec{p}(S') - \tau(S')$. We denote by $u(S) = \vec{v}(S) \cdot \vec{p}(S) - \tau(S)$ his expected utility for reporting his true type. According to Definitions 1 and 2, the incentive compatibility and individual rationality constraints are as follows.

Definition 18. *A direct mechanism is Incentive Compatible (IC) if for all $S, T \subseteq N$, $v(S) \cdot \vec{p}(S) - \tau(S) \geq \vec{v}(S) \cdot \vec{p}(T) - \tau(T)$, or equivalently, $u(S) \geq u(T) + (\vec{v}(S) - \vec{v}(T)) \cdot \vec{p}(T)$.*

Definition 19. *A direct mechanism is individually rational (IR) if for all $S \subseteq N$, $v(S) \cdot \vec{p}(S) - \tau(S) \geq 0$, or equivalently, $u(S) \geq 0$.*

To prove Theorem 8, we narrow into a family of instances for which there is a unique optimal IC, IR, direct mechanism, and this mechanism satisfies the following: for a special item i^* and a special type S^* , $p_{i^*}(S^*) \in \{0, 1\}$ but it is #P-hard to decide whether $p_{i^*}(S^*) = 0$. Our approach was outlined in Section 7.1.1 and is provided in detail in Sections 7.3 through 7.5.

Finally, it is straightforward to translate this hardness result to general mechanisms by using the ‘‘Revelation Principle’’ [Mye79, NRTV07]. This establishes Theorem 8 (for unrestricted mechanisms).

7.3 A Linear Programming Approach

Our goal in Sections 7.3 through 7.5 is showing that it is computationally hard to compute a IC, IR direct mechanism that maximizes the seller’s expected revenue, even in the single-bidder setting introduced in Section 7.2. In this section, we define three exponential-size linear programs which are useful for zooming into a family of hard instances that are also amenable to analysis. Our LPs are defined using the notation introduced in Section 7.2.

7.3.1 Mechanism Design as a Linear Program

The optimal IC and IR mechanism for the family of single-bidder instances introduced in Section 7.2 can be found by solving the following linear program, which we call LP1.

| |
|--|
| $\begin{array}{ll} \text{subject to :} & \max E_S[\vec{v}(S) \cdot \vec{p}(S) - u(S)] \\ \forall S, T \subseteq N : & u(S) \geq u(T) + (\vec{v}(S) - \vec{v}(T)) \cdot \vec{p}(T) \quad (\text{IC}) \\ \forall S \subseteq N : & u(S) \geq 0 \quad (\text{IR}) \\ \forall S \subseteq N, i \in N : & 0 \leq p_i(S) \leq 1 \quad (\text{PROB}) \end{array}$ |
|--|

Figure 7-1: LP1, the linear program for revenue maximization

Notice that the expression $\vec{v}(S) \cdot \vec{p}(S) - u(S)$ in the objective function equals the price $\tau(S)$ that the bidder of type S pays when reporting S to the mechanism. The expectation is taken over all $S \subseteq N$, where the probability of set S is given by $q(S) = \prod_{i \in S} q_i \cdot \prod_{j \notin S} (1 - q_j)$. We notice that this program has exponential size (variables and constraints).

7.3.2 A Relaxed Linear Program

We now remove constraints from LP1 and perform further simplifications, making the program easier to analyze. Later on we identify a subclass of instances where optimal solutions to the relaxed program induce optimal solutions to the original program (see Lemma 16).

As a first step, we relax LP1 by considering only IC constraints that correspond to neighboring types (types that differ in one element). That is, instead of requiring that no type of player can benefit by falsely declaring to be *any* other type, we only require that no type of player can benefit by falsely declaring to be a *neighboring* type. For simplicity in our analysis, we also drop the constraint that the probabilities $p_i(S)$ are non-negative:

$$\max E_S[\vec{v}(S) \cdot \vec{p}(S) - u(S)]$$

subject to :

$$\forall S \subseteq N, i \notin S : \quad u(S \cup \{i\}) \geq u(S) + d_i p_i(S) \quad (\text{IC1})$$

$$\forall S \subseteq N, i \notin S : \quad u(S) \geq u(S \cup \{i\}) - d_i p_i(S \cup \{i\}) \quad (\text{IC2})$$

$$\forall S \subseteq N : \quad u(S) \geq 0 \quad (\text{IR})$$

$$\forall S \subseteq N, i \in N : \quad p_i(S) \leq 1 \quad (\text{PROB'})$$

Since the coefficient of every $p_i(S)$ in the objective is strictly positive, no $p_i(S)$ can be increased in any optimal solution without violating a constraint. We therefore conclude the following about $p_i(S)$:

- If $i \in S$, then $p_i(S)$ is only upper-bounded by constraint PROB', and thus $p_i(S) = 1$ in every optimal solution.
- If $i \notin S$, then $p_i(S) = \min\{1, \frac{u(S \cup \{i\}) - u(S)}{d_i}\}$ from (IC1) and (PROB'). Furthermore, from (IC2) we have $\frac{u(S \cup \{i\}) - u(S)}{d_i} \leq p_i(S \cup \{i\}) = 1$, and thus $p_i(S) = \frac{u(S \cup \{i\}) - u(S)}{d_i}$.

So the program becomes (after setting $p_i(S) = 1$ whenever $i \in S$, removing the constant terms from the objective, and tightening the constraints (IC1) to equality):

$$\max E_S \left[\sum_{i \notin S} v_i(S) p_i(S) - u(S) \right]$$

subject to :

$$\forall S \subseteq N, i \notin S : \quad p_i(S) = \frac{u(S \cup \{i\}) - u(S)}{d_i} \quad (\text{IC1}')$$

$$\forall S \subseteq N, i \notin S : \quad u(S \cup \{i\}) - u(S) \leq d_i \quad (\text{IC2})$$

$$\forall S \subseteq N : \quad u(S) \geq 0 \quad (\text{IR})$$

$$\forall S \subseteq N, i \notin S : \quad p_i(S) \leq 1 \quad (\text{PROB}'')$$

Notice that the constraint (PROB'') is trivially satisfied as a consequence of (IC1') and (IC2). We now rewrite the objective, substituting $p_i(S)$ by (IC1') and noting that $v_i(S) = a_i$ for $i \notin S$:

$$E_S \left[\sum_{i \notin S} a_i \frac{u(S \cup \{i\}) - u(S)}{d_i} - u(S) \right] = E_S \left[u(S) \left(-1 - \sum_{i \notin S} \frac{a_i}{d_i} + \sum_{i \in S} \left(\frac{a_i}{d_i} \cdot \frac{q(S \setminus \{i\})}{q(S)} \right) \right) \right]$$

obtained by grouping together all coefficients of $u(S)$, adjusting by the appropriate probabilities. We note that $\frac{q(S \setminus \{i\})}{q(S)} = -1 + \frac{1}{q_i}$, and our objective becomes $E_S \left[u(S) \left(-1 - \sum_{i \in N} \frac{a_i}{d_i} + \sum_{i \in S} \frac{a_i}{q_i d_i} \right) \right]$.

We now perform a change of notation so that the program takes a simpler form. We set $B \leftarrow \kappa \left(1 + \sum_{i \in N} \frac{a_i}{d_i} \right)$ and $x_i \leftarrow \frac{\kappa a_i}{q_i d_i}$, where κ is some positive constant. The objective now becomes $\frac{1}{\kappa} E_S [(\sum_{i \in S} x_i - B)u(S)]$. Since $1/\kappa$ is constant, we study the following program, LP2:

$$\begin{aligned} & \max_u E_S [(\sum_{i \in S} x_i - B)u(S)] \\ \text{subject to :} & \\ \forall S \subseteq N, i \notin S : & \quad u(S \cup \{i\}) - u(S) \leq d_i \quad (\text{IC2}) \\ \forall S \subseteq N : & \quad u(S) \geq 0 \quad (\text{IR}) \end{aligned}$$

Figure 7-2: LP2, the relaxed linear program

In constructing LP2, our constants B and x were a function of \vec{q} , \vec{a} , \vec{d} , and a newly introduced constant κ . We note that, by adjusting κ , we can obtain a wide range of relevant B and x values.

Lemma 14. *For any B , \vec{x} , \vec{q} and \vec{d} such that $B > \sum_{i \in N} q_i x_i$, there exist (efficiently computable) \vec{a} and κ such that $B = \kappa(1 + \sum_{i \in N} \frac{a_i}{d_i})$ and $x_i = \frac{\kappa a_i}{q_i d_i}$. If B , \vec{x} , \vec{q} and \vec{d} are rational, then \vec{a} and κ are rational as well.*

Proof. We want that $x_i = \frac{\kappa a_i}{q_i d_i}$ and $B = \kappa + \sum_{i \in N} \frac{\kappa a_i}{d_i} = \kappa + \sum_{i \in N} q_i x_i$. Indeed, these equalities follow from setting $\kappa \leftarrow B - \sum_{i \in N} q_i x_i$ and $a_i \leftarrow \frac{q_i d_i x_i}{\kappa}$. \square

7.3.3 The Min-Cost Flow Dual of the Relaxed Program

To characterize the structure of optimal solutions to LP2, we use linear programming duality. Consider LP3, LP2's dual program, which has a (flow) variable $f_{S \cup \{i\} \rightarrow S}$ for every set S and $i \notin S$.

| |
|---|
| $\min_f \sum_S \sum_{i \notin S} f_{S \cup \{i\} \rightarrow S} d_i$ <p>subject to :</p> $\forall S \subseteq N : \quad - \sum_{i \notin S} f_{S \cup \{i\} \rightarrow S} + \sum_{i \in S} f_{S \rightarrow S \setminus \{i\}} \geq q(S) (\sum_{i \in S} x_i - B)$ $\forall S \subseteq N, i \notin S : \quad f_{S \cup \{i\} \rightarrow S} \geq 0$ |
|---|

Figure 7-3: LP3, the dual of LP2

We interpret LP3 as a minimum-cost flow problem on a lattice. Every node on the lattice corresponds to a set $S \subseteq N$, and flow may move downwards from S to $S \setminus \{i\}$ for each $i \in S$. The variable $f_{S \rightarrow S \setminus \{i\}}$ represents the amount of flow sent this way, and the cost of sending each unit of flow along this edge is d_i .

For nodes S with $q(S) (\sum_{i \in S} x_i - B) \geq 0$, we have an external source supplying the node with at least this amount of flow. We call such a node “positive.” Nodes with $q(S) (\sum_{i \in S} x_i - B) < 0$, which we call “negative,” can deposit at most $|q(S) (\sum_{i \in S} x_i - B)|$ to an external sink. Since $d_i > 0$ for all i , an optimal solution to LP3 will have net imbalance exactly $q(S) (\sum_{i \in S} x_i - B)$ for each positive node S .

7.4 Characterizing Linear Programming Solutions

For the remainder of this section, we restrict our attention to the case where N is the only positive node in LP3. We notice that there is a feasible solution if and only if $q(N) (\sum_{i \in N} x_i - B) \leq - \sum_{S \subsetneq N} q(S) (\sum_{i \in S} x_i - B)$, which occurs when $E_S [\sum_{i \in S} x_i - B] \leq 0$. Since the components of S are chosen independently, the program is feasible precisely when $\sum_{i \in N} q_i x_i - B \leq 0$.

7.4.1 The Canonical Solution to LP3

When there is a single positive node, we can easily construct an optimal solution to LP3 as follows. Define the cost of each node S to be $cost(S) \triangleq \sum_{i \in N \setminus S} d_i$, which corresponds to the cost of sending a unit of flow from N to S in LP3. (The flow can be sent along any path to the node, since all such paths have the same cost.) We order the negative nodes in increasing order of cost (and lexicographically if there are ties). We greedily send flow to the negative nodes in order, moving to the next node only when all previous nodes have been saturated. We stop when a net flow of $q(N) (\sum_{i \in N} x_i - B)$ has been absorbed by the negative nodes. We call this the *canonical solution* to LP3, and notice that the canonical solution is the unique optimal solution to LP3 up to the division of flow between equal cost nodes.

7.4.2 From LP3 to LP2 Solutions

We now show how to use a canonical solution to LP3 to construct a solution to LP2. In most instances, this solution is unique.

Lemma 15. *Let S^* be the highest-cost negative node which absorbs non-zero flow in the canonical solution f of LP3, and suppose that S^* is not fully saturated by f . Then the utility function $u(S) = \max\{cost(S^*) - cost(S), 0\}$ is the unique optimal solution to LP2.*

Proof. Consider an arbitrary optimal LP2 solution u . We will use linear programming complementarity to prove that u is uniquely determined by the canonical solution f .

For any node S that receives nonzero flow in f , there is a path $N = S_0, S_1, S_2, \dots, S_k = S$ from N to S that has positive flow along each edge. By complementarity, the (IC2) inequalities corresponding to these edges in the primal program are tight in u . That is, for all $i = 1, \dots, k$, we have $u(S_{i-1}) - u(S_i) = d_x$, where x is the unique element of $S_{i-1} \setminus S_i$. So, for any S which receives nonzero flow in f :

$$u(N) - u(S) = \sum_{i \in N \setminus S} d_i = cost(S).$$

For all nodes S' which are not fully saturated in f (i.e. S^* as well as all nodes which receive no flow), $u(S')$ must be 0 in u by complementarity, since the corresponding LP3 constraints are not tight. In particular, since S^* receives flow but is not fully saturated, we have $u(S^*) = 0$ and hence:

$$u(N) = u(N) - u(S^*) = \text{cost}(S^*).$$

Therefore, any node S which receives flow in f must have $u(S) = u(N) - \text{cost}(S) = \text{cost}(S^*) - \text{cost}(S)$.

Furthermore, a node S always receives flow in f if its cost is less than $\text{cost}(S^*)$, and receives no flow if its cost is greater than $\text{cost}(S^*)$. Moreover, those nodes S with $\text{cost}(S) = \text{cost}(S^*)$ either receive no flow in which case $u(S) = 0$, or receive flow in which case $u(S) = \text{cost}(S^*) - \text{cost}(S) = 0$. Thus, we have shown that for any node S , $u(S) = \max\{\text{cost}(S^*) - \text{cost}(S), 0\}$. It is easy to verify that this utility function satisfies all the constraints of LP2. \square

If the highest cost node S^* to receive flow in f were fully saturated, then the utility function described in Lemma 15 would still be an optimal LP2 solution. However, in this case, if the cheapest unfilled node in f had strictly greater cost than S^* , then the optimal primal solution would not be unique.

7.4.3 From LP2 to LP1 Solutions

We now show that, in certain cases, a solution to LP2 allows us to obtain a solution to LP1 where \vec{q} , \vec{a} and \vec{d} are as in Lemma 14. The proof of Lemma 16 sets p as in Section 7.3.2 and then verifies that (u, p) satisfies all the constraints of LP1.

Lemma 16. *Suppose $B > \sum_{i \in N} q_i x_i$ and an optimal solution u to LP2 is monotone and supermodular. Then there is some p such that (u, p) is an optimal solution to LP1 where \vec{q} , \vec{a} , \vec{d} are as in Lemma 14. If u is the unique optimal solution to LP2, then (u, p) is the unique optimal LP1 solution.*

Proof. We set

$$p_i(S) = \begin{cases} 1, & \text{if } i \in S; \\ \frac{u(S \cup \{i\}) - u(S)}{d_i}, & \text{otherwise.} \end{cases}$$

With this choice, as explained in Section 7.3.2, (u, p) is an optimal solution to a relaxation of LP1. So to establish optimality of (u, p) for LP1 it suffices to show that (u, p) satisfies all the constraints of LP1.

We first notice that the (IR) constraints are satisfied, since $u(S) \geq 0$ for all S in LP2.

We now show that the (PROB) constraints are satisfied. Indeed, if $i \in S$, then $p_i(S) = 1$. If $i \notin S$, then $p_i(S) \geq 0$ follows from monotonicity of u . The inequality $p_i(S) \leq 1$ follows from constraint (IC2) of LP2.

Finally, we show that the (IC) constraints of LP1 are satisfied. By supermodularity of u we have that for all S , all $i \notin S$ and all $j \neq i$:

$$u(S \cup \{i\} \cup \{j\}) - u(S \cup \{j\}) \geq u(S \cup \{i\}) - u(S).$$

Dividing by d_i we obtain $p_i(S \cup \{j\}) \geq p_i(S)$ for all $i \notin S$ and $j \neq i$. Since the inequality is trivially satisfied if $i \in S$ (since both sides are 1), or $j = i$ (since $p_i(S \cup \{i\}) = 1$) we conclude that \vec{p} is monotone.

Now pick any distinct subsets $S, T \subseteq N$. We must show that:

$$u(S) \geq u(T) + (\vec{v}(S) - \vec{v}(T)) \cdot \vec{p}(T).$$

Consider an ordering i_1, i_2, \dots, i_k of the elements of $T \setminus S$ and an ordering j_1, j_2, \dots, j_ℓ of the elements of $S \setminus T$.

By (IC2), we know that, for all $r = 1, \dots, k$:

$$u\left(S \cup \bigcup_{t=1}^r \{i_t\}\right) \leq u\left(S \cup \bigcup_{t=1}^{r-1} \{i_t\}\right) + d_{i_r}.$$

Summing over r and canceling terms, we conclude $u(S \cup T) \leq u(S) + \sum_{r=1}^k d_{i_r}$.

From our definition of \vec{p} it follows that for all $r = 1, \dots, \ell$:

$$u\left(T \cup \bigcup_{t=1}^r \{j_t\}\right) = u\left(T \cup \bigcup_{t=1}^{r-1} \{j_t\}\right) + d_{j_r} p_{j_r}\left(T \cup \bigcup_{t=1}^{r-1} \{j_t\}\right).$$

By monotonicity of \vec{p} , it follows that

$$u\left(T \cup \bigcup_{t=1}^r \{j_t\}\right) \geq u\left(T \cup \bigcup_{t=1}^{r-1} \{j_t\}\right) + d_{j_r} p_{j_r}(T).$$

Summing over r , we conclude that $u(S \cup T) \geq u(T) + \sum_{r=1}^{\ell} d_{j_r} p_{j_r}(T)$.

Combining this with our earlier upper bound for $u(S \cup T)$, we conclude that

$$u(S) \geq u(T) + \sum_{r=1}^{\ell} d_{j_r} p_{j_r}(T) - \sum_{r=1}^k d_{i_r}.$$

Since $p_{i_r}(T) = 1$ for all r , we have

$$u(S) \geq u(T) + \sum_{j \in S \setminus T} d_j p_j(T) - \sum_{i \in T \setminus S} d_i p_i(T)$$

and thus the (IC) constraint of LP1 is satisfied.

If u is the unique optimal solution to LP2, then the (u, p) constructed as above is the unique optimal solution to LP1, as it is the unique optimal solution of a relaxation of LP1. \square

7.4.4 Putting it All Together

In summary, we have shown that if the canonical solution of LP3 has a partially saturated node S^* , then LP2 has a unique optimal solution, namely $u(S) = \max\{\text{cost}(S^*) - \text{cost}(S), 0\}$. Since this utility function is monotone and supermodular, it also defines a unique optimal solution of the corresponding LP1 instance.

Corollary 2. *Let S^* be the highest-cost negative node which absorbs non-zero flow in the canonical solution of LP3, and suppose that S^* is not fully saturated. Then the original mechanism design problem with \vec{q} , \vec{a} and \vec{d} as in Lemma 14 has a unique*

optimal solution, and the utility of a player of type N in this solution is $\text{cost}(S^*)$.

7.5 Proof of Theorem 8

We use the results of the previous section to establish the computational hardness of optimal mechanism design. Our reduction is from the lexicographic rank problem, which we show to be #P-hard.

Definition 20 (LEXRANK problem). *Given a collection $\mathfrak{C} = \{c_1, \dots, c_n\}$ of positive integers and a subset $S \subseteq \{1, \dots, n\}$, we define the lexicographic rank of S , denoted $\text{lexr}_{\mathfrak{C}}(S)$, by*

$$\left| \left\{ S' : |S'| = |S| \text{ and } \left(\sum_{i \in S'} c_i < \sum_{j \in S} c_j \text{ or } \left(\sum_{i \in S'} c_i = \sum_{j \in S} c_j \text{ and } S' \leq_{\text{lex}} S \right) \right) \right\} \right|$$

where $S' \leq_{\text{lex}} S$ is with respect to the lexicographic ordering.⁶ The LEXRANK problem is: Given \mathfrak{C} , S , and an integer k , determine whether or not $\text{lexr}_{\mathfrak{C}}(S) \leq k$.

Lemma 17. LEXRANK is #P-hard.

Proof. We show that the LEXRANK problem is #P-hard by a reduction from #-SUBSETSUM.

Definition 21 (#-SUBSETSUM problem). *Given a collection $\mathcal{W} = \{w_1, \dots, w_n\}$ of positive integers and a target integer T , compute the number of subsets $S \subseteq \{1, \dots, n\}$ such that $\sum_{i \in S} w_i \leq T$.*

The #-SUBSETSUM problem is known to be #P-hard. Indeed, the reduction from SAT to SUBSETSUM as presented in [Sip06] is parsimonious.

Given an oracle for the LEXRANK problem, it is straightforward to do binary search to compute the lexicographic rank of a set S . We will prove hardness of LEXRANK by reducing the #-SUBSETSUM problem to the computation of lexicographic ranks of a collection of sets.

⁶To be precise, we say that $S_1 \leq_{\text{lex}} S_2$ iff the largest element in the symmetric difference $S_1 \Delta S_2$ belongs to S_2 .

Let (\mathcal{W}, T) be an instance of $\#$ -SUBSETSUM, where $\mathcal{W} = \{w_1, \dots, w_n\}$ is a collection of positive integers and T is a target integer. We begin by defining, for $m = 1, \dots, n$:

$$\text{count}_{\mathcal{W}}(T, m) \triangleq \left| \left\{ S \subseteq \{1, \dots, n\} : |S| = m \text{ and } \sum_{i \in S} t_i \leq T \right\} \right|.$$

Note that the number of subsets of \mathcal{W} which sum to at most T is simply $\sum_{m=1}^n \text{count}_{\mathcal{W}}(T, m)$.

So it suffices to compute $\text{count}_{\mathcal{W}}(T, m)$ for all m .

To do this, we define n different collections $\mathfrak{C}_1, \dots, \mathfrak{C}_n$, where $\mathfrak{C}_\ell = \{c_1^\ell, \dots, c_{n+\ell}^\ell\}$ is given by:

$$c_i^\ell = \begin{cases} 4nw_i & \text{if } 1 \leq i \leq n \\ 4nT + 2n & \text{if } i = n + 1 \\ 1 & \text{if } n + 2 \leq i \leq n + \ell. \end{cases}$$

We also define a special set $S_\ell \triangleq \{n + 1, n + 2, \dots, n + \ell\}$. Notice that $\sum_{i \in S_\ell} c_i^\ell = 4nT + 2n + \ell - 1$. Furthermore, for every subset $S \subseteq \{1, \dots, n\}$, we have

$$\sum_{i \in S} c_i^\ell = 4n \sum_{i \in S} w_i.$$

Hence, for all $\emptyset \neq S \subseteq \{1, \dots, n\}$:

1. if $\sum_{i \in S} w_i > T$, then $\sum_{i \in S} c_i^\ell > \sum_{j \in S_\ell} c_j^\ell$;
2. if $\sum_{i \in S} w_i \leq T$, then for all $U \subseteq \{n+2, n+3, \dots, n+\ell\}$ we have that $\sum_{i \in S \cup U} c_i^\ell < \sum_{j \in S_\ell} c_j^\ell$;
3. for all $U \subseteq \{n + 2, n + 3, \dots, n + \ell\}$, $\sum_{i \in S \cup U \cup \{n+1\}} c_i^\ell > \sum_{j \in S_\ell} c_j^\ell$.

Given that $|S_\ell| = \ell$ the above imply

$$\text{lex}_{\mathfrak{C}_\ell}(S_\ell) = 1 + \sum_{m=1}^{\ell} \left(\text{count}_{\mathcal{W}}(T, m) \cdot \binom{\ell - 1}{\ell - m} \right).$$

Suppose we have an oracle which can compute the lexicographic rank of a given set. We first use this oracle to determine $\text{lexr}_{\mathfrak{C}_1}(S_1)$ and thereby compute $\text{count}_{\mathcal{W}}(T, 1)$. Next, we use the oracle to determine $\text{lexr}_{\mathfrak{C}_2}(S_2)$ and thereby compute $\text{count}_{\mathcal{W}}(T, 2)$, using the previously computed value of $\text{count}_{\mathcal{W}}(T, 1)$. Continuing this procedure n times, we can compute $\text{count}_{\mathcal{W}}(T, m)$ for all $m = 1, \dots, n$. This concludes the proof. \square

We will now give a reduction from LEXRANK to OMD. Let (\mathfrak{C}, S, k) be an instance of LEXRANK where $\mathfrak{C} = \{c_1, \dots, c_n\}$ is a collection of positive integers, $S \subseteq \{1, \dots, n\}$, and k is integer. We wish to determine whether $\text{lexr}_{\mathfrak{C}}(S) \leq k$. We assume that $|S| \neq 0, n$ as otherwise the problem is trivial to solve.

We denote by $[n]$ the set $\{1, 2, \dots, n\}$ and $[n + 1] = [n] \cup \{n + 1\}$. We construct an OMD instance indirectly, by defining an instance of LP3 with the following parameters:

- $d_i = 2^{n+1} \left(c_i + \sum_{j=1}^n c_j \right) + 2^i$, for $i \in [n]$ and $d_{n+1} = 1$;
- $x_i = 2$, for all i ;
- $B = 2n + 1$;
- $q_i = q$ for all i , where we leave $q \in [0.5, 1 - \frac{1}{2n+2})$ a parameter.

We note that $B > \sum_i x_i q_i$, and thus Lemma 14 implies that, for all q , an instance of LP3 as above arises from some OMD instance $\{a_i, d_i, q_i\}_{i=1}^{n+1}$, in the notation of Section 7.2.

Denote by S^c the set $[n] \setminus S$.⁷ Suppose that, for some value q , there is a partially filled node T^* in the canonical LP3 solution such that $T^* \subseteq [n]$ and $|T^*| = n - |S|$. Using Lemma 15 we have

$$\begin{aligned} p_{n+1}(S^c) &= \frac{u^*(S^c \cup \{n+1\}) - u^*(S^c)}{1} \\ &= \max\{\text{cost}(T^*) - \text{cost}(S^c \cup \{n+1\}), 0\} - \max\{\text{cost}(T^*) - \text{cost}(S^c), 0\} \\ &= \max\{\text{cost}(T^*) - \text{cost}(S^c) + 1, 0\} - \max\{\text{cost}(T^*) - \text{cost}(S^c), 0\} \end{aligned}$$

⁷Note that $\{n+1\}$ is in neither S nor S^c .

Therefore, since the cost of each set is an integer, $p_{n+1}(S^c) = \begin{cases} 0 & \text{if } \text{cost}(S^c) > \text{cost}(T^*) \\ 1 & \text{if } \text{cost}(S^c) \leq \text{cost}(T^*) \end{cases}$.

Since $n + 1$ is in neither S^c nor T^* , $p_{n+1}(S^c) = \begin{cases} 0 & \text{if } \sum_{i \in S} d_i > \sum_{j \in [n] \setminus T^*} d_j \\ 1 & \text{if } \sum_{i \in S} d_i \leq \sum_{j \in [n] \setminus T^*} d_j \end{cases}$. By our construction of the d_i 's we can see that since $|T^*| = n - |S|$,

$$p_{n+1}(S^c) = \begin{cases} 0 & \text{if } \left\{ \sum_{i \in S} c_i > \sum_{j \in [n] \setminus T^*} c_j \right\} \text{ or } \left\{ \sum_{i \in S} c_i = \sum_{j \in [n] \setminus T^*} c_j \text{ and } S >_{\text{lex}} ([n] \setminus T^*) \right\} \\ 1 & \text{if } \left\{ \sum_{i \in S} c_i < \sum_{j \in [n] \setminus T^*} c_j \right\} \text{ or } \left\{ \sum_{i \in S} c_i = \sum_{j \in [n] \setminus T^*} c_j \text{ and } S \leq_{\text{lex}} ([n] \setminus T^*) \right\}. \end{cases}$$

Therefore, $p_{n+1}(S^c) = 1$ if $\text{lexr}_{\mathfrak{C}}(S) \leq \text{lexr}_{\mathfrak{C}}([n] \setminus T^*)$ and 0 otherwise.

So our next goal is to set the parameter q such that there is a partially filled node T^* in the canonical LP3 solution such that $T^* \subseteq [n]$, $|T^*| = n - |S|$, and $\text{lexr}_{\mathfrak{C}}([n] \setminus T^*) = k$. For such q , distinguishing between $p_{n+1}(S^c) = 0$ and $p_{n+1}(S^c) = 1$ would allow us to solve the LEXRANK instance. The next lemma shows that a q as required can be found in polynomial time.

Lemma 18. *In polynomial time, we can identify a $\tilde{q} \in [0.5, 1 - \frac{1}{2n+2})$ with $O(n \log n)$ bits of precision such that the partially filled node in the canonical LP3 solution with parameter $q = \tilde{q}$ is a set $T^* \subseteq [n]$ of size $n - |S|$ and $\text{lexr}_{\mathfrak{C}}([n] \setminus T^*) = k$.*

Proof. In our construction for the proof of Theorem 8, the lowest cost negative node is $[n]$. Furthermore, the cost of every negative node is unique, and for any $T \subsetneq [n]$ there is no node with cost between that of $T \cup \{n + 1\}$ and T . Also, if T and T' are proper subsets of $[n]$ and if $|T| > |T'|$, then $\text{cost}(T) < \text{cost}(T')$.

For each i between 1 and $n - 1$, let T_1^i, T_2^i, \dots be the ordering of the size- i subsets of $[n]$ in increasing order of cost. In the canonical LP3 solution, $[n]$ fills first, and T_j^i fills before $T_{j'}^{i'}$ if it has larger size ($i > i'$) or the same size but smaller cost ($i = i'$ and $j < j'$). Furthermore, each node $T_j^i \cup \{n + 1\}$ fills immediately before the node T_j^i .

Our goal is to choose q so that $T_k^{n-|S|}$ is partially filled. Indeed, the sets $[n] \setminus T_1^{n-|S|}$ through $[n] \setminus T_k^{n-|S|}$ are precisely the sets counted in the computation of $\text{lexr}_{\mathfrak{C}}([n] \setminus T_k^{n-|S|})$. Notice that lexicographic tie-breaking of lexr is enforced by construction of

adding an additional 2^i to d_i .

The only positive node, $[n+1]$, emits a net flow of q^{n+1} , and the node $[n]$ absorbs at most $q^n(1-q)$ flow. For each size i between $n-1$ and $n-|S|+1$, there are $\binom{n}{i}$ sets $T \subseteq [n]$ of size i , each of which can absorb

$$|q(T)(2|T| - B)| = (2(n-i) + 1)q^i(1-q)^{n+1-i}$$

flow. Furthermore, each set $T \cup \{n+1\}$ can absorb

$$|q(T \cup \{n+1\})(2|T| + 2 - B)| = (2(n-i) - 1)q^{i+1}(1-q)^{n-i}.$$

Thus, in total, T and $T \cup \{n+1\}$ can absorb

$$\begin{aligned} & q^i(1-q)^{n-i} ((1-q)(2n-2i+1) + q(2n-2i-1)) \\ &= q^i(1-q)^{n-i}(2(n-i-q) + 1). \end{aligned}$$

Finally, we notice that T is responsible for at least a $1/(2n+2)$ fraction of the quantity above, since

$$\frac{(2(n-i) + 1)q^i(1-q)^{n+1-i}}{(2(n-i-q) + 1)q^i(1-q)^{n-i}} \geq 1 - q > \frac{1}{2n+2},$$

using that $q < 1 - \frac{1}{2n+2}$.

If all nodes strictly preceding (i.e. with smaller cost than) $T_1^{n-|S|} \cup \{n+1\}$ have been saturated, the amount of flow still unabsorbed is

$$\begin{aligned} & q^{n+1} - q^n(1-q) - \sum_{i=n-|S|+1}^{n-1} \binom{n}{i} q^i(1-q)^{n-i}(2(n-i-q) + 1) \\ &= \sum_{i=n-|S|+1}^n \binom{n}{i} q^i(1-q)^{n-i}(2(i+q-n) - 1). \end{aligned}$$

Therefore, a sufficient condition for $T_k^{n-|S|}$ to be partially filled in the canonical solu-

tion is

$$f(q) \triangleq \frac{\sum_{i=n-|S|+1}^n \binom{n}{i} q^i (1-q)^{n-i} (2(i+q-n) - 1)}{q^{n-|S|} (1-q)^{|S|} (2|S| - 2p + 1)} \in \left(k - \frac{1}{2n+2}, k \right).$$

We claim that there is such a $q^* \in [0.5, 1 - \frac{1}{2n+2})$ such that $f(q^*) = k - \frac{1}{4n+4}$. Indeed, for $q < 0.5$, only $[n]$ will ever receive flow, so in this case, $f(q) < 0$. Furthermore, we can lower-bound $f(q)$ by the following ratio (where we add a negative quantity to the numerator)

$$\begin{aligned} & \frac{\sum_{i=n-|S|+1}^n \binom{n}{i} q^i (1-q)^{n-i} (2(i+q-n) - 1) + \sum_{i=0}^{n-|S|-1} \binom{n}{i} q^{n-i} (1-q)^i (2(i+q-n) - 1)}{q^{n-|S|} (1-q)^{|S|} (2|S| - 2p + 1)} \\ &= \frac{\sum_{i=0}^n \binom{n}{i} q^i (1-q)^{n-i} (2(i+q-n) - 1) - \binom{n}{|S|} q^{n-|S|} (1-q)^{|S|} (-2|S| + 2p - 1)}{q^{n-|S|} (1-q)^{|S|} (2|S| - 2p + 1)} \end{aligned}$$

and thus

$$\begin{aligned} f(q) &\geq \binom{n}{|S|} - \frac{\sum_{i=0}^n \binom{n}{i} q^i (1-q)^{n-i} (2n - 2i - 2p + 1)}{q^{n-|S|} (1-q)^{|S|} (2|S| - 2p + 1)} \\ &= \binom{n}{|S|} - \frac{2n - 2p + 1 - 2 \sum_{i=0}^n i \binom{n}{i} q^i (1-q)^{n-i}}{q^{n-|S|} (1-q)^{|S|} (2|S| - 2p + 1)} \\ &= \binom{n}{|S|} - \frac{2n - 2p + 1 - 2pn}{q^{n-|S|} (1-q)^{|S|} (2|S| - 2p + 1)}. \end{aligned}$$

Hence, for $q = 1 - \frac{1}{2n+2}$, we get $f(q) \geq \binom{n}{|S|} \geq k$. Using this, the continuity of f , and that $f(q) < 0$ for $q < 0.5$, we conclude that there is a $q^* \in [0.5, 1 - \frac{1}{2n+2})$ such that $f(q^*) = k - \frac{1}{4n+4}$.

We now consider $\tilde{q} = q^* \pm \epsilon \in [0.5, 1 - \frac{1}{2n+2})$. We claim that $f(\tilde{q}) \in (k - \frac{1}{2n+2}, k)$ as long as $\epsilon = O\left(\frac{(4n)^{-4n}}{4n+4}\right)$. To show this, we bound the absolute value of the derivative $\frac{df}{dq}$ at all points in $[0.5, 1 - \frac{1}{2n+2})$. The numerator of $\frac{df}{dq}$ is a polynomial in q of degree $2n+1$, where the coefficient of each term is, in absolute value, $O(2^{3n} \cdot \text{poly}(n)) \leq O(2^{4n})$ —using the crude bound $\binom{n}{i} \leq 2^n$. Furthermore, the denominator $(q^{n-|S|} (1-q)^{|S|} (2|S| - 2p + 1))^2$ of $\frac{df}{dq}$ is greater than $(\frac{1}{2n+2})^{2n}$, since $q \in [0.5, 1 - \frac{1}{2n+2})$. Therefore, we can bound the magnitude of the derivative by $O((2n+2)^{2n} (2n+1) 2^{4n}) = O((4n)^{4n})$.

Since this bound holds for all points in $[0.5, 1 - \frac{1}{2n+2})$, we conclude that

$$f(\tilde{q}) \in (f(q^*) - \epsilon O((4n)^{4n}), f(q^*) + \epsilon O((4n)^{4n}))$$

and thus, for $\epsilon = O\left(\frac{(4n)^{-4n}}{4n+4}\right)$ we have that $f(\tilde{q}) \in (k - \frac{1}{2n+2}, k)$.

Thus, we can find the desired \tilde{q} in polynomial time via binary search on $f(q)$, requiring $O(n \log n)$ bits of precision in q . \square

We conclude our reduction for the proof of Theorem 8 as follows: First, we compute \tilde{q} as in Lemma 18. Next, we solve the OMD instance resulting from this choice of parameter. The solution allows us to sample in expected polynomial-time from the allocation and price rule of the optimal auction for type S^c . We have proven that this type receives item $n + 1$ with probability 1 if $(\mathfrak{C}, S, k) \in \text{LEXRANK}$ and with probability 0 otherwise. A single sample from the allocation rule of the optimal auction suffices to tell which one is the case, since if $p_{n+1}(S^c) = 1$ then every sample from the allocation rule should allocate the item and if $p_{n+1}(S^c) = 0$ then no sample should allocate the item. So if we can solve OMD in expected polynomial-time, then we can solve LEXRANK in expected polynomial-time by a. finding \tilde{q} using Lemma 18; b. solving the resulting OMD instance; c. drawing a single sample from the allocation rule of the auction for type S^c and outputting “yes” if and only if item $n + 1$ is allocated by the drawn sample.

7.6 Beyond Additive or Computationally Bounded Bidders

7.6.1 Budget-Additive Bidders

We have already remarked in Section 2.2 that when the bidder valuations have combinatorial structure, it becomes much easier to embed computationally hard problems into the optimal mechanism design problem. In this spirit, Dobzinski et al. [DFK11] show that optimal mechanism design for OXS bidders is NP-hard. While we won't

define OXS valuations, we illustrate the richness in their structure by recalling that in [DFK11] the items are taken to be edges of a graph $G = (V, E)$, and there is a single bidder whose valuation is drawn from a distribution that includes in its support the following valuation:

$$f(E') = \max \left\{ |A| \mid \begin{array}{l} A \subseteq E', \text{ every connected component of} \\ G' = (V, A) \text{ is either acyclic or unicyclic} \end{array} \right\}, \forall E' \subseteq E.$$

While the valuations used in [DFK11] are quite rich, we notice here that the techniques of [DFK11] can be used to establish hardness of optimal mechanism design for a bidder with very mild combinatorial structure, namely budget-additivity. Indeed, consider the problem of selling a set $N = \{1, \dots, n\}$ of items to a budget-additive bidder whose value x_i for each item is some deterministically known integer, but whose budget is only probabilistically known. In particular, suppose that, with probability $(1 - \epsilon)$, the bidder is additive, in which case his value $v_a(S)$ for each subset S is $\sum_{i \in S} x_i$. With probability ϵ , however, he has a positive integer budget B : he values each subset S at $v_b(S) = \min\{\sum_{i \in S} x_i, B\}$. That is, when he has a budget, he receives at most B utility from any subset. We claim that the optimal mechanism satisfies:

Claim 9. *Suppose $\epsilon < \frac{1}{1 + \sum x_i}$. Then every optimal individually rational and Bayesian incentive compatible direct mechanism for the budget-additive bidder described above has the following form:*

- *If the bidder is unbudgeted, he receives all items and is charged $\sum_{i \in N} x_i$.*
- *If the bidder is budgeted, he receives a probability distribution over the subsets T of items such that $\sum_{i \in T} x_i$ is as large a value as possible without exceeding B . he is charged that value.*

This claim follows from a lemma of [DFK11], showing that if the bidder has no budget he must receive his value maximizing bundle, while if he is budgeted he must receive some bundle T maximizing $v_b(T) - (1 - \epsilon)v_a(T)$. Determining the optimal direct mechanism is clearly hard in this context, as it is NP-hard to compute a subset T with the largest value that does not exceed B . Using the same arguments as in

Section 7.2 we can extend this lower bound to all mechanisms, by noticing that any sample from the allocation to the budgeted bidder answers whether there exists some T such that $\sum_{i \in T} x_i = B$.

Theorem 9. *There is a polynomial-time Karp reduction from the subset-sum problem to the optimal mechanism design problem of selling multiple items to a single budget-additive quasi-linear bidder whose values for the items are known rational numbers, and whose budget is equal to some finite rational value or $+\infty$ with rational probabilities.*

7.6.2 Further Discussion: Powerful Bidders

NP Power

We observe that if the budget-additive bidder described in the previous section has the ability to solve NP-hard problems, then there is a simple indirect mechanism which implements the above allocation rule: The seller offers each set S at price $\sum_{i \in S} x_i$, and leaves the computation of the set T to the bidder. (As remarked in Section 7.2 we can use small rewards to guarantee that the same allocation rule is implemented.) Such a mechanism is intuitively unsatisfactory, however, and violates our requirement (see Section 7.1) that a bidder be able to compute his strategy efficiently. While an optimal indirect mechanism is easy for the seller to construct and to implement, it is intractable for the bidder to determine his optimal strategy in such a mechanism. Conversely, implementing the direct mechanism requires the seller to solve a subset sum instance. Thus, shifting from a direct to an indirect mechanism allows for a shift of computational burden from the seller to the bidder.

PSPACE and Beyond

Indeed, we can generalize this observation to show that optimal mechanism design for a polynomial-time seller becomes much easier if the bidder is quasi-linear, computationally unbounded, and has a known finite maximum possible valuation for an

allocation. The intuition is that, if some canonical optimal direct mechanism is computable and implementable in polynomial space, then we can construct an extensive form indirect mechanism whereby the bidder first declares his type and then convinces the seller, via an interactive proof with completeness 1 and low soundness, what the allocation and price would be for his type in the canonical optimal direct revelation IC and IR mechanism.⁸ In the event of a failure, the bidder is charged a large fine, significantly greater than his maximum possible valuation of any subset. The proof that this protocol achieves, in a subgame perfect equilibrium, expected revenue equal to the optimal direct IC and IR revenue follows from $IP = PSPACE$ [Sha92]. In particular, we note that the canonical solution of the mechanism design instances in the hardness proof of Section 7.5 can indeed be found in $PSPACE$, and thus we can construct an easily-implementable extensive form game which achieves optimal revenue in subgame perfect equilibrium. Reaching this equilibrium, however, requires the bidder to have $PSPACE$ power.

Indeed, if there are two or more computationally-unbounded bidders, then we can replace the assumption that the optimal direct mechanism is computable in $PSPACE$ with the even weaker assumption that it is computable in nondeterministic exponential time, using the result that $MIP = NEXP$ [BFL91] to achieve the optimal revenue in a perfect Bayesian equilibrium. The intuition is that our mechanism now asks all players to (simultaneously) declare their type. Two players are selected and required to declare the allocation and price which the canonical optimal direct mechanism would implement given this type profile, and they then perform a multiparty interactive proof to convince the seller of the correctness of the named allocation and price.

Since the optimal allocation might be randomized and the prices might require exponentially many bits to specify, we must modify the above outline. The rough ideas are the following. First, a price of exponential bit complexity but magnitude bounded by a given value can be obtained as the expectation of a distribution that

⁸We have fixed a canonical direct mechanism to avoid complications arising when the optimal mechanism is not unique.

can be sampled with polynomially many random bits in expectation. Now, after the bidders declare their types, the seller reveals a short random seed r , and requires the bidders to draw the prices and allocation from the appropriate distribution, using r as a seed. The bidders must prove the actual deterministic allocation and prices that the optimal direct mechanism would have obtained when given seed r to sample exactly from the optimal distribution. If r does not have sufficient bits to determine the sample, then the bidders prove interactively that additional bits are needed, and further random bits are appended to r . If the bidders ever fail in their proof, the mechanism terminates and they are charged a large fine. Regardless of whether or not the players ever attempt to prove an incorrect statement, the protocol terminates in expected polynomial time.

We omit a formal proof of the above observations. We also note that we do not present the mechanisms of this section as practical. But we want to point out that the complexity of the optimal mechanism design problem may become trivial if no computational assumptions are placed on the bidders.

Part II

Two-Sided Information Asymmetry

Chapter 8

Overview and Related Work

Myerson’s work on single item auctions as well as our extension to multi-item auctions assume a private values model, where bidders know exactly how much the item for sale is worth to them. The uncertainty of the seller in the bidders’ valuations is modeled in a Bayesian manner, assuming that a distribution from which the bidders’ valuations are drawn is known to all — the auctioneer and all bidders. If independence of the private valuations is assumed, a remarkably elegant mathematical analysis yields a revenue-optimal auction for single-item auctions.

In contrast with the standard Bayesian model with private values, in several applications such as ad auctions the value of each bidder (advertiser) for an item (user) may depend on several attributes of the item that are unobservable to him. In ad auctions, it depends in intriguing ways on the advertiser’s precise business model, but also on detailed information about the user that is available, typically through cookies, to the auctioneer; that is to say, there is significant information asymmetry in ad auctions. For example, it is typical that the value of an advertiser changes significantly depending on the age, the location or the gender of a user. This is more pronounced in display ads than in search ads, since in the latter the auctioneer and the bidder do share at least one important piece of information: the keyword.

Such situations with two sided information asymmetry are captured by a more general auction model introduced by Milgrom and Weber [[MW82](#)]. They assume that the auctioneer as well as each agent receive real-valued signals, which together with

some extra signals not observed by anyone determine the value of each agent for the item. Everyone is assumed to have consistent priors about all signals that he does not observe, and each bidder's value is the same, non-decreasing function of his own signal, the set of all other bidders' signals, and the signals observed by the auctioneer and by nobody. The signals are also subtly correlated, in that their joint distribution is assumed to be *affiliated*, satisfying an FKG-type technical condition. Intuitively, this condition ensures that larger values for some signals render more likely larger values for other signals. Hence, a larger value for one agent renders more likely larger values for the others.

Milgrom and Weber analyze in this model the three common auction types, and order them with respect to expected auctioneer revenue: The English auction fares best, followed by the second-price auction, with the first-price auction trailing; reserve prices and entry fees are also considered. Then they prove that in all these auctions, *the auctioneer can increase the expected revenue by adopting a policy of always revealing all available information, i.e. the signals he observes and any other affiliated signal he can obtain.* This remarkable result, compatible with age old practices by auction houses, is known as the *linkage principle*, mathematical articulation of the intuition that more information means that all bidders are subject to less risk and information costs, and can therefore be a little bolder in their bidding.

Note that Milgrom and Weber do not address the *combined problem*: What is the optimum *augmented auction*? By “augmented auction” we mean here a design that specifies both an information revelation policy and an allocation policy (potentially multi-round) to be used by the auctioneer. In optimization, it is well known that separately optimizing two aspects of a design can result in gross suboptimalities. Hence, finding the optimum augmented auction requires an optimal co-design of the information revelation and the allocation rule of the auction. The motivation of Milgrom and Weber was different. They aimed to understand the role of information revelation in the most commonly used single-item auction formats, and they established linkage in a general model of affiliation. In contrast, we aim to identify the optimal augmented auction: Here we consider a simple model of bidder valuations explicitly

motivated by ad auctions and identify the optimum augmented auction. Importantly, our results apply also to settings without affiliation, involving multiple items, and not single-dimensional bidders.

We proceed to explain our model and results, after providing some additional context. Information revelation in auctions through signaling has reemerged in the study of ad auctions in three recent papers [EFG⁺14, BMS12, FJM⁺12]. In [EFG⁺14, BMS12], each bidder's valuation is a function of the his own type and the item's type, both types being discrete. The assumption is that the auctioneer knows the realized item type, each bidder knows his own type, and everyone knows the joint distribution of types of all bidders and the item. Given this distribution, the authors seek to optimize the signaling by the auctioneer preceding a second-price auction, and conclude that designing the optimal signaling scheme is computationally highly non-trivial in general. Similarly, in [FJM⁺12], every bidder has a valuation function which depends on the item type and his own type. Importantly, the bidder type is assumed scalar (i.e. the bidder is single-dimensional) and the valuation function is increasing in the buyer's type. Each bidder knows his own type, the auctioneer knows the item types, and all types are assumed independent from distributions known to everybody. Following the formulations of [EFG⁺14, BMS12], which have their roots in the signaling literature, they consider signaling schemes where the auctioneer commits to sending some (potentially randomized) signal $\sigma(t_i)$ drawn from a distribution $Q(t_i) \in \Delta(\Sigma)$ that depends on the realized item type t_i , where Σ is some signal alphabet. Upon receiving the signal, the bidders are expected to update their beliefs about the item type. Hence, the auctioneer should run Myerson's optimal auction for the distributions reflecting the bidders' updated expectations for the item value, conditioning on the signal. For this two-stage format, they show that full revelation of information by the auctioneer, namely using $\sigma(t_i) = t_i$, is optimal, i.e. linkage holds. It is quite subtle and surprising this mechanism (of optimal signaling followed by Myerson's optimal auction) is a suboptimal solution of the signaling and auction co-design problem that we consider here, even for a single bidder, as the following example suggests.

Example 6. We have a single bidder and a single item. There are two possible item types, H and L , and three possible bidder types A , B and C ; both item and bidder types are uniformly random. The bidder’s value for the item as a function of the item and the bidder types is shown in this table:

| | | Item Types | |
|--------------|-----|------------|-----|
| | | H | L |
| Bidder Types | A | 6 | 6 |
| | B | 10 | 6 |
| | C | 20 | 16 |

Notice that the valuation defined by the table is non-decreasing in the bidder’s type, for each item type. Hence, as implied by [FJM⁺ 12], the optimal signaling followed by Myerson’s auction is to perfectly reveal the item type and charge 10 for item H and 6 for item L . The resulting revenue is $6 + \frac{1}{3} \approx 6.333$. In this auction, we “go after” types B and C for item H , and all three types for item L . In particular, type C always has a surplus of at least 10. How can we get some of that surplus? Can we target C , without losing revenue from types A and B ?

Here is how: Before revealing the item type, we announce the prices (10 for H and 6 for L) and then ask the bidder either to choose now one of the two item types foregoing his right to buy the other, or pay 5 for the privilege of postponing this choice until after the item type is revealed. Then we reveal the item type, and sell it — provided the bidder has not lost his right to buy it. It is easy to see that, in this protocol, a bidder of type A will waive his right to H , and will therefore pay 3 in expectation. A bidder of type B will waive his right to L , thereby paying 5 in expectation. However, a bidder of type C will prefer to pay 5 upfront, and then pay another 8 in expectation, for a total of 13. Notice that even though the new auction decreased the expected payment of type B from 8 to 5, the expected payment of type C increased from 8 to 13. This auction, by deviating from [FJM⁺ 12]’s optimal signaling followed by Myerson’s optimal auction, managed to “target” a particular bidder type

and thus achieve a better expected revenue of 7¹. Therefore, full information disclosure and the linkage principle may not hold — even when the distributions are affiliated — when followed by auction formats other than the four classical ones considered by Milgrom and Weber. This concludes Example 6.

In further recent work on signaling, [DIR14] study the optimum signaling scheme to precede a second-price auction under restrictions on the signals that can be sent, while [CCD⁺15] develop algorithmic tools for optimum signaling under general objectives and constraints. Besides the works on signaling in auctions outlined so far, there are several other papers addressing the problems of “matching” and “targeting” [Boa09, AABG13, GMMV15, BHV14, HM15], variants of the information disclosure conundrum, in various models which yield interesting results. Linkage—that is, advantage of disclosure by the auctioneer—sometimes prevails, under assumptions, and sometimes does not. None of the papers in the literature solves or considers what we call “the augmented auction design problem,” the problem of co-designing the optimal signaling and allocation rule.

Finally, an interesting precursor of our result is the work in [BP07]. They study a problem conceptually quite close to ours involving the simultaneous optimization over both information structure and allocation policy, in a setting where the bidder has no private information and the seller can directly reveal to the bidder what the value is *without observing that value*; interestingly, the optimal combined mechanism involves a coarse revelation followed by a Myerson auction. [ES07] and [BW15]) extend that model to allow private information to the bidders and study more general sequential auctions. In contrast, in our model multiple rounds of interaction cannot improve revenue (Theorem 10).

¹An equivalent implementation of this auction is through *discount*: Announce the prices to be 20 and 16 and give the bidder a 10-off discount coupon to be applied to an item type of his choice prior to the realization of the item type.

8.1 Our Model and Results

In our model, an auctioneer is selling a single item,² which may be of n possible types and whose realized type is drawn from a publicly known distribution $\vec{\pi}$. There are also m bidders interested in the item, who receive types independently from publicly known distributions D_1, \dots, D_m . Each distribution D_i is an n -dimensional distribution. Every vector in the support of D_i is a possible type for bidder i and specifies the bidder's value for each type of item. An important special case of our problem is when there is a single bidder, i.e. $m = 1$. In this case, the setting can be summarized in a table V_{ij} , whose rows and columns correspond respectively to possible types of the bidder and the item, together with a distribution over rows (bidder types) and a distribution over columns (item types). Let us consider an example.

Example 7. *There are two item types, uniformly distributed in $\{\text{red}, \text{blue}\}$, and one buyer with 25 possible types, uniformly distributed on the set $\{0, 1, 2, 3, 4\} \times \{0, 1, 2, 3, 4\}$. In Section 10.1.2, we show that the optimal signaling-followed-by-Myerson auction achieves revenue 1.2. In fact, as we show in Section 10.1.1, just the plain old Myerson auction without any information revelation achieves revenue of 1.2. Nevertheless, as we show in Section 10.1.3, the optimal revenue is 1.28. One way to achieve this revenue is to give the bidder the following options, before revealing the item type:*

- *Pay 1.5 upfront for the option of getting the item if it is blue;*
- *Pay 1.5 upfront for the option of getting the item if it is red;*
- *Pay 2.5 upfront to get the item independently of its color.*

It is easy to see that a buyer of types $(3, 0), (3, 1), (4, 0)$ or $(4, 1)$ will purchase the “red” option, and similarly for the “blue” option. In contrast, a buyer of type $(2, 3)$,

²All our results extend to multi-item settings, as discussed in Chapter 12, but we focus our discussion to single-item settings for simplicity of the exposition. Moreover, our results extend to settings where there is correlation between the item type and the bidder types

$(2, 4)$, $(3, 2)$, $(3, 3)$, $(3, 4)$, $(4, 2)$, $(4, 3)$ or $(4, 4)$ will for the last option, paying 2.5. Hence, the revenue equals $\frac{8}{25} \cdot 1.5 + \frac{8}{25} \cdot 2.5 = 1.28$.

Another way to get the optimal revenue does not even require the auctioneer to learn the item type. Here is the auction: The auctioneer offers the buyer two options:

- Pay 2.5 to get the item;
- Pay 3 to get the item, with the additional option of returning it for full refund.

It is easy to check that the types who went after the third option in the previous implementation, will go for the first option now. Intuitively, these types are strong and they are seeking a “discount.” The weaker types, who are more skewed and went for the first two options in the previous implementation, will go after the second option now, as it represents “insurance.” This insurance stratagem for increasing revenue is reminiscent of the role of warranties for avoiding collapse in Akerlof’s work on informationally problematic markets [Ake70].

The optimal augmented auctions obtained in Examples 6 and 7 with their nonstandard design principles — targeting, discounts, insurance, etc. — are manifestations of our main theorem, relating optimal augmented auction design with the optimal multi-item mechanism design problem in the setting with private values we presented in the first part.

Informal Theorem 1 (See Theorem 10). *The optimal augmented auction design for a single item of n possible types and m bidders is equivalent to an optimal multi-item auction design with n items and m additive bidders. Moreover, there is an implementation of the optimal augmented auction in which the auctioneer reveals no information about the item type to the bidder prior to bidding.*

Thus our main contribution is the observation that the augmented auction problem (finding the best co-design of signaling and auction format) is an inherently multi-dimensional problem, and all the phenomena/pathologies/nonstandard design principles identified in the examples above are manifestations of this multi-dimensional nature of the problem. The intuition for our theorem is simple: an auctioneer who

auctions a pen that is 50% red and 50% blue, as per Example 7, is in fact auctioning, to additive bidders, two items: One item that is a blue pen with probability 50% and nothing otherwise; and another item that is a red pen with probability 50% and nothing otherwise. Hence, the optimum solution to the augmented auction problem maps to the optimum two-item auction with additive bidders. If in the latter some bidder pays p for the (fractional) red pen, this means that in the former the same bidder pays p for the option of getting the red pen, if the pen is red. Importantly, the optimum augmented auction has an implementation where the auctioneer reveals no information about the item type to the bidder. That is, *when both auction format and signaling are optimized simultaneously, the optimal signaling is the null signaling.* Hence, the linkage principle does not prevail. Stronger than null signaling, sometimes the optimal augmented auction can be turned into a *no peeking auction*: in Example 7, the auctioneer can derive maximum revenue from the item without ever finding out if it is blue or red.

We also show that the above statements on auction optimality are more than academic curiosities: We can leverage the connection to multi-item auctions with additive bidders to show that the optimum augmented auction may result in dramatically more revenue than the suboptimal auction paradigms currently in use. Theorem 11 fleshes out the connections between augmented auctions and multi-item auctions, relating different signaling-auction formats to well-studied multi-item auction formats. Pushing the connection to the multi-item setting even further, Tables 11.1 and 11.2 provide a revenue comparison between the following auction formats for single and multi-bidder settings respectively and for different kinds of item/bidder type distributions: (a) running Myerson’s auction with no information revelation by the auctioneer, (b) running Myerson’s auction preceded by the optimal information revelation by the auctioneer, and (c) running the optimal augmented auction. We show that the approximation gaps between (c) and (b) may be unbounded.

One could object that we have not solved the problem of optimal augmented auctions, but instead we have *reduced it* to multi-item auctions with additive bidders, an important open problem in auction theory, notorious for its analytical difficulty. This

is all true, and yet our reduction has many and potent consequences. First, it is important to know the nature of the true optimum, to which an auction designer should aspire, because this way one understands better how existing solutions are in reality compromises to the true goal, and one can start exploring ways to improve. Second, there is significant recent progress in multi-item auctions. In Chapter 2 we outlined several computational and structural results for multi-item auctions. Moreover, our framework, developed in the first part, provides a characterization of single-bidder instances. Hence our reduction opens new avenues for improving on the status quo.

Another possible objection is that even the one-bidder solution of the multi-dimensional problem is in general quite complicated and of a stochastic nature, and therefore awkward to use in practice (the bidder chooses among a possibly long menu of complex lotteries involving subsets of the items). However, one can interpret the one-bidder optimal solution as *a set of possible contracts* offered by the auction designer, as an alternative to the auction. In a situation where there are millions of items of each type, and thousands of bidders of each type — not an unrealistic assumption for a day of ad auctions, assuming some coarse enough clustering of both sides — the probabilities of both kinds of types, as well as the probabilities in the lotteries, are multiplied by large enough numbers and are realized as magnitudes that are reasonably accurately known a priori. Such *large market interpretation* of the one-bidder solution may provide a new basis for understanding how ad contracts should be designed.

Chapter 9

Problem Formulation

9.1 A General Auction Model

We consider a general auction model similar to Milgrom and Weber [MW82]. In this setting, m interested bidders compete for the possession of a single item for sale¹.

Each bidder possesses some information about the item for sale. The information that bidder i possesses about the item is captured by a type variable t_i . Moreover, there are several quality features about the item that affect the bidders' values and are only observable by the seller. These features are captured by a type variable τ_{item} for the item that can be one of n different possibilities.

The actual value of bidder i , which depends on variables not observed by him, is given by a function $V_i(\tau_{item}, t_1, \dots, t_m)$ which is publicly known. The vector of types $(\tau_{item}, t_1, \dots, t_m)$ is drawn by a joint distribution D .

The seller is running an auction for the item seeking to maximize his expected revenue. The seller commits to a mechanism $\mathcal{M} = (\mathcal{P}, \mathcal{T})$ prior to observing the item type t_{item} . After every bidder i chooses a strategy $A_i(t_i)$ the mechanism interacts with all bidders revealing information about the type of the item and other bidders and asks them to take actions. After the choice of actions $\vec{A}(\vec{t}) = (A_1(t_1), \dots, A_m(t_m))$ and the interactions of agents with the mechanism the (possibly random) outcome $\mathcal{M}(\vec{A}) = (\mathcal{P}(\vec{A}), \mathcal{T}(\vec{A}))$ is implemented, where $\mathcal{P}_i(\vec{A})$ is an indicator variable indicating whether

¹Multiple items can be dealt with similarly, see the discussion in Chapter 12.

bidder i got the item and $\mathcal{T}_i(\vec{A})$ expresses how much the bidder had to pay in total. Thus, every bidder gets utility $u_i(\mathcal{M}(\vec{A}), \tau_{item}, \vec{t}) = \mathcal{P}_i(\vec{A})V_i(\tau_{item}, \vec{t}) - \mathcal{T}_i(\vec{A})$.

We are interested in two common implementation concepts of auctions, implementation in *ex-post Nash equilibria* and *Bayes-Nash equilibria*. In particular, we want to design auctions where bidders' strategies form either an ex-post or a Bayes-Nash equilibrium.

Definition 22. A strategy profile $(A_1(\cdot), \dots, A_m(\cdot))$ is an ex-post Nash equilibrium if

$$\mathbb{E}[u_i(\mathcal{M}(\vec{A}(\vec{t})), \tau_{item}, \vec{t})] \geq \mathbb{E}[u_i(\mathcal{M}(a_i, \vec{A}_{-i}(\vec{t}_{-i})), \tau_{item}, \vec{t})]$$

for all i, a_i and \vec{t} , where the expectation is over the type of the item $\tau_{item} \sim D|\vec{t}$ and the randomness in the choices of the mechanism.

Definition 23. A strategy profile $(A_1(\cdot), \dots, A_m(\cdot))$ is a Bayes Nash equilibrium if

$$\mathbb{E}[u_i(\mathcal{M}(\vec{A}(\vec{t})), \tau_{item}, \vec{t})] \geq \mathbb{E}[u_i(\mathcal{M}(a_i, \vec{A}_{-i}(\vec{t}_{-i})), \tau_{item}, \vec{t})]$$

for all i, a_i and t_i , where the expectation is over the type of the item $(\tau_{item}, \vec{t}_{-i}) \sim D|t_i$ and the randomness in the choices of the mechanism.

9.2 An Important Special Case

We restrict our attention to an interesting special case of the general model where the types of the item and all bidders, $\tau_{item}, t_1, \dots, t_m$, are drawn from independent distributions. Moreover, the value of every bidder i is only a function of his own type and the type of the item τ_{item} , i.e. $v_i(t_i, \tau_{item})$. Thus, the value of bidder i is captured by a vector $\vec{v}_i \in \mathbb{R}^n$, where v_{ij} gives the value of the bidder for item type $j \in [n]$. Thus, in this simpler model we can assume that the type of the buyer is given by a vector of values for every item.

We denote by $\vec{\pi}$ the prior probability distribution on item types and by D_i the distribution on n -dimensional vectors that gives the type of bidder i .

While our main result applies to the general model presented above as discussed in Chapter 12, we will state many of our results for this simpler model instead. This interesting special case has been studied by several recent papers [EFG⁺14, BMS12, FJM⁺12] with the goal to identify the amount of information to reveal about the item type that optimizes the seller’s revenue. This question is studied in the context of signaling where the seller might send signals to the bidder’s containing information about the item type and causing them to update their prior. Our results suggest that significantly higher revenue is achievable under the optimal augmented auction as opposed to running a standard single item auction that is prefaced by a round of signaling even in this simpler model with independent types.

9.3 Signaling Schemes

The seller may commit to a policy of revealing, prior to the auction, certain information about the item type to the buyers with the goal of increasing revenue. A *signaling scheme* $Q : [n] \rightarrow \Delta(S)$ maps each item type to a distribution over a set of signals S . Upon receiving a signal $s \in S$, each buyer updates his prior for the item $\vec{\pi}$ to a distribution $\vec{\Pi}(s)$, where

$$\vec{\Pi}_j(s) = \frac{\text{Prob}[\text{signal is } s | \text{item is } j] \cdot \text{Prob}[\text{item is } j]}{\text{Prob}[\text{signal is } s]} = \frac{Q_s(j)\pi_j}{\sum_{j'} Q_s(j')\pi_{j'}}.$$

Therefore, each signal $s \in S$ gives an updated prior $\vec{\Pi}(s)$. Given a collection of vectors $\vec{\Pi}(s)$ and their corresponding probabilities $P(s) = \text{Prob}[\text{signal is } s] = \sum_{j'} Q_s(j')\pi_{j'}$, they satisfy $\sum_{s \in S} P(s)\vec{\Pi}(s) = \vec{\pi}$.

Conversely, given a collection of vectors $\vec{\Pi}(s)$ with corresponding probabilities $P(s)$, there exists a signaling scheme Q producing these belief vectors with these probabilities if the above equality holds. This is easy to see by setting $Q_j(s) = P(s) \frac{\vec{\Pi}_j(s)}{\pi_j}$.

We can thus write an alternative parametrization of a signaling scheme in terms of $\vec{\Pi}$ and P .

9.4 Auction Formats

We consider three different auction formats depending on how restricted the seller is in the mechanism design process.

Myerson (No Information Revelation) This is the most restricted setting possible where the seller does not have any information about the item type or he is not allowed to reveal any information to the buyers and he is forced to run a standard single item auction. In this case that no information is revealed, every buyer i with value vector $\vec{v}_i \sim D_i$ has value $\vec{v}_i \cdot \vec{\pi}$ for the item and the setting becomes single dimensional. The standard Myerson auction gives us the revenue-optimal mechanism.

Myerson with Signaling This is the setting considered by Fu et al. [FJM⁺12] where the seller is restricted to run a Myerson auction but can observe the item type and may engage in signaling, which may cause the buyers to update their prior about the item.

Optimal Augmented Auction In the most flexible setting of all, the seller can observe the item type, reveal some information to the bidders and then ask them to take actions. This process might repeat for several rounds until a winner for the item is declared. The optimal augmented auction is the auction that results in the maximum expected revenue in the best ex-post or Bayes-Nash equilibrium.

We note that augmented auctions are significantly more general than Myerson with Signaling auctions. In particular, the optimal mechanisms in Examples 6 and 7 are augmented auctions that are not expressible in the Myerson-with-Signaling auction format.

Chapter 10

Handling Bidder Uncertainty

10.1 Optimal Auctions

10.1.1 Myerson (No Information Revelation)

When no information is available about the item type, the setting becomes one dimensional with bidder value distributions $\vec{\pi} \cdot D_i$. By $\vec{\pi} \cdot D_i$ we denote the distribution obtained by first sampling a vector $\vec{v}_i \sim D_i$ and then computing the inner product with the prior belief vector $\vec{\pi}$. From Fact 1, Myerson's auction achieves the highest revenue in this one dimensional setting which we denote by $\text{MYERSON}(\vec{\pi})$.

Recall example 7, where there are two item types ($n = 2$) and one buyer ($m = 1$) with D_1 being the uniform distribution over $\{0, 1, 2, 3, 4\} \times \{0, 1, 2, 3, 4\}$. The belief about the item type is uniform, i.e. $\vec{\pi} = (0.5, 0.5)$. If no information about the item type is revealed, the buyer can be collapsed to single-dimensional with value distribution $\vec{\pi} \cdot D_1$, which takes values in the set $\{0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4\}$. The optimal mechanism offers the item at a price of 2 which is accepted with probability 60%. This gives a revenue of 1.20.

10.1.2 Myerson with Signaling

Let us denote by $\text{MSIGNALING}(\vec{\pi})$ the maximum revenue the seller can achieve by fixing a signaling scheme and then running Myerson, where beliefs $\vec{\Pi}(s)$ occur with

probability $P(s)$, i.e.:

$$\text{MSIGNALING}(\vec{\pi}) = \max_{(\vec{\Pi}, P)} \sum_s P(s) \cdot \text{MYERSON}(\vec{\Pi}(s))$$

where the maximum is computed over signaling schemes such that $\sum_{s \in S} P(s) \vec{\Pi}(s) = \vec{\pi}$.

We can show that, without loss of generality, we can restrict ourselves to signaling schemes with at most n signals (this is a fairly common phenomenon in the signaling literature).

Lemma 19. *Consider a signaling scheme $(\vec{\Pi}, P)$. There exists a signaling scheme $(\vec{\Pi}', P')$ that achieves at least as much revenue in the Myerson with Signaling setting which uses at most n signals.*

Proof. By reverse induction. Assume there exists a signaling scheme with $|S| > n$ signals. We will construct a signaling scheme with $|S| - 1$ signals achieving the same revenue. Consider the vectors $\vec{\Pi}(s)$ for all $s \in S$. Since $|S| > n$ the vectors are not linearly independent. Thus, there exist constants α_s which are not all 0 such that $\sum_s \alpha_s \vec{\Pi}(s) = \vec{0}$. Splitting the summation terms into positive and negative we get that $\sum_{s: \alpha_s > 0} |\alpha_s| \vec{\Pi}(s) = \sum_{s: \alpha_s < 0} |\alpha_s| \vec{\Pi}(s)$.

Now consider the corresponding difference in the revenue between the positive and negative sides, i.e. $\sum_s \alpha_s \text{MYERSON}(\vec{\Pi}(s))$. Without loss of generality, we can assume that this quantity is non-negative, since otherwise we can pick a new vector $\vec{\alpha}' = -\vec{\alpha}$. Furthermore, let $\lambda = \min_{s: \alpha_s < 0} \left(\frac{P(s)}{|\alpha_s|} \right)$ and s^* be the corresponding index where this is minimized.

We create a new signaling scheme with corresponding probabilities $P'(s) = P(s) + \lambda \alpha_s$. This makes the generated revenue at least as much as it was before:

$$\begin{aligned} \sum_s (P(s) + \lambda \alpha_s) \text{MYERSON}(\vec{\Pi}(s)) &\geq \sum_s P(s) \text{MYERSON}(\vec{\Pi}(s)) + \lambda \sum_s \alpha_s \text{MYERSON}(\vec{\Pi}(s)) \\ &\geq \sum_s P(s) \text{MYERSON}(\vec{\Pi}(s)) \end{aligned}$$

Moreover, $P'(s^*) = 0$ since $P(s^*) + \frac{P(s^*)}{|\alpha_{s^*}|} \alpha_{s^*} = 0$ which means that the new signaling scheme uses at most $|S| - 1$ signals. By induction, we conclude that repeating this process we can obtain a signaling scheme that uses at most n signals which achieves at least as much revenue as the original signaling scheme. \square

We can also show an upper bound on the revenue of the optimal Myerson auction with signaling as follows:

Lemma 20. *The maximum revenue achievable in the Myerson with Signaling setting is at most*

$$\text{MSIGNALING}(\vec{\pi}) \leq \max_{\vec{p} \geq 0: |\vec{p}|=1} \text{MYERSON}(\vec{p})$$

Moreover, the above holds with equality if all distributions are item-symmetric and the prior probabilities $\vec{\pi}$ are uniform.

Proof. For any signaling scheme $(\vec{\Pi}, P)$ it holds that:

$$\begin{aligned} \text{MSIGNALING}(\vec{\pi}) &= \max_{(\vec{\Pi}, P)} \sum_s P(s) \text{MYERSON}(\vec{\Pi}(s)) \leq \max_{(\vec{\Pi}, P)} \max_s \text{MYERSON}(\vec{\Pi}(s)) \\ &\leq \max_{\vec{p} \geq 0: |\vec{p}|=1} \text{MYERSON}(\vec{p}) \end{aligned}$$

Now, let \vec{p}^* be the vector that maximizes the right-hand side. In an item-symmetric setting with uniform prior probabilities on the item types, we can construct a feasible signaling scheme $(\vec{\Pi}, P)$ by considering all cyclic rotations of the vector \vec{p}^* with equal probability. Since, all distributions are item symmetric, all cyclic rotations give revenue as $\text{MYERSON}(\vec{p}^*)$. This proves that in the symmetric setting, the inequality shown above is tight. \square

We revisit Example 7. Applying Lemma 20, we get the optimal revenue obtainable in the Myerson with signaling setting is upper bounded by

$$\max_{0 \leq p \leq 1} \text{MYERSON}(p, 1 - p)$$

which is maximized without loss of generality for some

$$p \in \left\{ \frac{a}{b} \mid 0 \leq a \leq b \leq 4 \text{ for } a, b \in \mathbb{Z} \right\}$$

Checking each of these values separately we see that the revenue is at most 1.2. This upper bound can be achieved by always revealing the item type to the buyer. Once the item type is revealed, the buyer's value is known to be distributed uniformly in $\{0, 1, 2, 3, 4\}$ and the optimal auction is to set a single price p that maximizes the expected revenue against this distribution. This is achieved by selling the item at 2 which is accepted with 60% probability. The same revenue can also be achieved by giving no information as in the previous section. We conclude that both signaling schemes are optimal.

10.1.3 Optimal Augmented Auction

We now study the problem of designing the combination of signaling scheme and subsequent auction that maximizes revenue and is either ex-post or Bayes-Nash implementable. We show that this is equivalent to the following multidimensional mechanism design problem:

Definition 24 (Optimal Multi-Item Auction). *We have m additive bidders and n items. Bidder i 's value for item j is equal to $v_{ij}\pi_j$ where the vector \vec{v}_i comes from bidder distribution D_i and $\vec{\pi}$ is a common parameter for all agents. We denote by $\text{REV}(\vec{\pi})$ the optimal revenue achievable by any incentive-compatible mechanism, given the parameter $\vec{\pi}$.*

Theorem 10. *The optimal augmented auction for a given $\vec{\pi}$ and D_1, \dots, D_m is equivalent to a multi-item auction with n items and m additive bidders as specified in Definition 24. In particular, for any ex-post implementable augmented auction, there exists an incentive-compatible multi-item auction that achieves the same revenue. Similarly, for any Bayes-Nash implementable augmented auction, there exists a Bayesian incentive-compatible multi-item auction that achieves the same revenue.*

Moreover, there is an implementation of the optimal augmented auction in which the auctioneer reveals no information about the item type to the bidder prior to bidding.

The proof of the theorem is quite simple and follows (i) by observing that the revelation principle holds in settings with two-sided information asymmetry, and (ii) by characterizing the auctions resulting from the revelation principle. What is subtle and noteworthy here is that the optimum is not a Myerson-with-signaling auction, even in the case of a single-dimensional bidder considered in [FJM⁺12]. Still, the optimum auction can be expressed succinctly through a one-to-one correspondence with a related multi-item auction with additive bidders.

Proof. Consider a mechanism \mathcal{M} that is ex-post implementable. The mechanism \mathcal{M} may use multiple rounds of communication and information revelation to the buyers. For each $i \in [m]$, let $A_i(\vec{v}_i)$ be the (possibly randomized) equilibrium strategy of buyer i when his type is \vec{v}_i , i.e. his value for every item type j is v_{ij} . Each strategy $A_i(\vec{v}_i)$ specifies the actions that bidder i takes in the mechanism \mathcal{M} given information about the item type and the actions of other buyers, in order to maximize buyer i 's expected payoff.

Let $\mathcal{P}_i(j, \vec{A})$ be an indicator random variable that indicates whether buyer i gets the item when buyers choose strategies $\vec{A} = (A_1, \dots, A_m)$ and the realized item type is $j \in [n]$. Similarly, let $\mathcal{T}_i(j, \vec{A})$ denote the price buyer i is asked to pay. The expected utility of buyer i is then equal to

$$\mathbb{E}_{j \sim \pi} [\mathbb{E}[\mathcal{P}_i(j, \vec{A})v_{ij} - \mathcal{T}_i(j, \vec{A})]]$$

where the first expectation is with respect to the randomness of the item type, while the second expectation is with respect to the randomness in the choices of the mechanism, the information revealed and the actions of the buyers under the mixed strategies \vec{A} .

We now examine the equilibrium conditions for the buyer strategies. For all possible types $\vec{V} = (\vec{v}_1, \dots, \vec{v}_m)$ and all possible misreports \vec{v}'_i for buyer i , we let

$\vec{A}(\vec{V}) = (A_1(\vec{v}_1), \dots, A_m(\vec{v}_m))$ and $\vec{A}(\vec{v}'_i, \vec{V}_{-i}) = (A_1(\vec{v}_1), \dots, A_i(\vec{v}'_i), \dots, A_m(\vec{v}_m))$. The equilibrium conditions for the strategy profile $\vec{A}(\vec{V})$ imply that:

$$\begin{aligned} \mathbb{E}_{j \sim \pi}[\mathbb{E}[\mathcal{P}_i(j, \vec{A}(\vec{V}))v_{ij} - \mathcal{T}_i(j, \vec{A}(\vec{V}))]] &\geq \\ \mathbb{E}_{j \sim \pi}[\mathbb{E}[\mathcal{P}_i(j, \vec{A}(\vec{v}'_i, \vec{V}_{-i}))v_{ij} - \mathcal{T}_i(j, \vec{A}(\vec{v}'_i, \vec{V}_{-i}))]] & \end{aligned}$$

Now set the variables $x_{ij}(\vec{V}) = \mathbb{E}[\mathcal{P}_i(j, \vec{A}(\vec{V}))]$ and $c_i(\vec{V}) = \mathbb{E}_{j \sim \pi}[\mathbb{E}[\mathcal{T}_i(j, \vec{A}(\vec{V}))]]$. The equation above can be rewritten for all i, \vec{V} and potential misreports v'_i as:

$$\sum_{j=1}^n x_{ij}(\vec{V})\pi_j v_{ij} - c_i(\vec{V}) \geq \sum_{j=1}^n x_{ij}(\vec{v}'_i, \vec{V}_{-i})\pi_j v_{ij} - c_i(\vec{v}'_i, \vec{V}_{-i}) \quad (\text{IC})$$

Moreover, the mechanism must satisfy the individual rationality constraint in order for the bidders to participate

$$\sum_{j=1}^n x_{ij}(\vec{V})\pi_j v_{ij} - c_i(\vec{V}) \geq 0 \quad (\text{IR})$$

and that for all $j \in [n]$: $\sum_{i=1}^m x_{ij}(\vec{V}) \leq 1$. A mechanism that satisfies all these constraints and achieves expected revenue

$$\mathbb{E}_{\vec{V} \sim D_1 \times \dots \times D_m} \left[\sum_{i=1}^m c_i(\vec{V}) \right]$$

can be viewed as a multi-item direct auction for m additive bidders with n items where item j is simply π_j units of item type j as in Definition 24. The direct mechanism allocates the item j to bidder i with probability $x_{ij}(\vec{V})$ when all bidders have types \vec{V} . The (IC) constraint implies that the auction is incentive compatible while the (IR) constraint ensures bidder participation. Any direct mechanism \mathcal{M} for the multi-item setting is a feasible augmented auction since the seller can use \mathcal{M} to presell all item types before observing the real item type and then decide on the final allocation after the type is observed.

The same reduction goes through for converting any Bayes-Nash equilibrium to

an equivalent Bayesian incentive-compatible direct mechanism that presells all the items. The difference is that in the (IC) and (IR) constraints above for bidder i , we take an expectation over all bidder types \vec{V}_{-i} other than i . \square

Therefore, the optimal augmented auction achieves revenue $\text{REV}(\vec{\pi})$, as defined in Definition 24, when the prior on the items is $\vec{\pi}$.

Remark 8. One way to convince somebody that no information revelation before running the optimal augmented auction can possibly increase revenue is to notice that the function $\text{REV}(\cdot)$ is concave.

Lemma 21. *The optimal revenue function $\text{REV}(\vec{x})$, as per Definition 24, is concave.*

Proof. We must have that for all $x, y \in \mathbb{R}_+^n$, $\frac{1}{2}\text{REV}(\vec{x}) + \frac{1}{2}\text{REV}(\vec{y}) \leq \text{REV}\left(\frac{\vec{x} + \vec{y}}{2}\right)$. Since the revenue function is homogeneous, this is equivalent to showing that $\text{REV}(\vec{x}) + \text{REV}(\vec{y}) \leq \text{REV}(\vec{x} + \vec{y})$. Let M_x and M_y be the optimal mechanisms that achieve $\text{REV}(\vec{x})$ and $\text{REV}(\vec{y})$. We can show that $\text{REV}(\vec{x} + \vec{y})$ achieves higher revenue by constructing a mechanism that sells to the buyers a fraction of the items using mechanism M_x and the rest using M_y . In particular, in M_x , we sell $\frac{x_j}{x_j + y_j}$ units of every item j while in M_y , we sell $\frac{y_j}{x_j + y_j}$. This achieves revenue $\text{REV}(\vec{x}) + \text{REV}(\vec{y})$. \square

So what if we sent a signal and then run an optimal augmented auction for the posterior of the item type? Suppose $(\vec{\Pi}, P)$ is a signaling scheme, as per Section 9.3. This signaling followed by the optimal augmented auction would result in revenue $\sum_s P(s)\text{REV}(\vec{\Pi}(s)) \leq \text{REV}(\sum_s P(s)\vec{\Pi}(s)) = \text{REV}(\vec{\pi})$. The inequality follows from Jensen's inequality and Lemma 21. This confirms that no signaling scheme can preface running the optimal augmented auction on the posterior to increase revenue (which of course we knew already from optimality, but the concavity illustrates one reason why it happens). This concludes Remark 8. \blacksquare

We presented our Theorem for the special case where type distributions for the bidders and the item are independent and every bidder's value is not affected by the types of the others. In Chapter 12, we provide a proof for the general model, where we

show that the two-sided information asymmetry setting is equivalent to a multi-item auction with interdependent values.

We now revisit Example 7 to compute the optimal augmented auction. It follows from Theorem 10 that this can be computed by solving an associated multi-item auction problem. The result¹ is the auction that gives the following choices to the buyer:²

- Buy one item at price 1.5.
- Buy both items at price 2.5.

We can see that if the buyer has values (3, 0), (3, 1), (4, 0) or (4, 1), he purchases only item 1 and similarly for item 2. The bundle of both items is bought by types (2, 3), (2, 4), (3, 2), (3, 3), (3, 4), (4, 2), (4, 3) and (4, 4). This means that the revenue is equal to $\frac{8}{25} \cdot 1.5 + \frac{8}{25} \cdot 2.5 = 1.28$. This is strictly better than the revenue achievable by running Myerson with signaling.

10.2 Connections to Multi-Item Mechanisms

In Theorem 10, we showed that the combined problem of designing a signaling scheme together with an optimal mechanism to sell the item to the bidders is equivalent to the multi-item mechanism design problem in Definition 24. In particular, we showed that no signaling is required and that the seller can treat the mechanism design problem as a multi-item auction where all types of items are available but their values are discounted according to the prior probability vector. Then, it becomes clear that the seller should not reveal any information but instead run an auction preselling each of the item types. The auction format should be the one that maximizes revenue in the equivalent multi-item setting.

¹The optimality of the mechanism was verified by solving a small linear program that encodes all the incentive compatibility constraints. For an analytical derivation in a similar setting with item values distributed uniformly in [0,1], we refer the reader to Section 4.3.1.

²We state the solution to the associated multi-item problem here. In Chapter 8, we have already described what this auction corresponds to for the optimal augmented problem, and gave two possible implementations of the optimal augmented auction.

It is interesting to consider how the restricted formats of auctions we studied above translate to the multi-item setting.

We first look at the case where no information is given to the buyers and a simple Myerson auction is run for the item. In this setting, the item is given to the winner of the auction no matter what type it is. This is equivalent to a multi-item mechanism for the setting of Definition 24 that bundles all items together and sells the bundle as one. This bundling mechanism is a well-studied simple mechanism in the multi-item auction literature, denoted as BREV. It performs quite well compared to the optimal mechanism in cases where the bidder values for the bundle of all items are well-concentrated. One important case where this happens is when items are identical and the value distributions for every item are identical and independent. We consider such settings in Chapter 11 where we make a detailed comparison of how well the optimal augmented auction is approximated by simple signaling schemes.

We now consider the case of running a simple Myerson auction but using signaling to increase revenue. In this restricted auction format, we can see that the optimal revenue can be written as

$$\begin{aligned} \text{MSIGNALING}(\vec{\pi}) &= \max_{(\vec{\Pi}, P)} \sum_s P(s) \cdot \text{MYERSON}(\vec{\Pi}(s)) = \max_{(\vec{\Pi}, P)} \sum_s \text{MYERSON}(P(s) \cdot \vec{\Pi}(s)) \\ &= \max_{\substack{\vec{x}_1 + \dots + \vec{x}_n = \vec{\pi} \\ \vec{x}_j \in \mathbb{R}_{\geq 0}^n}} \sum_{j=1}^n \text{MYERSON}(\vec{x}_j) \end{aligned} \tag{10.1}$$

This partitions the items into at most n groups according to the vectors \vec{x}_j . In the full information setting where the item type is always revealed to the buyers, we have that $\vec{x}_j = \pi_j \vec{e}_j$, where $\vec{e}_1, \dots, \vec{e}_n$ are the standard basis vectors. When viewed as a multi-item auction as in Definition 24, this corresponds to selling each item separately, by running separate Myerson auctions. This is an important mechanism, denoted by SREV, which is studied in the multi-item literature since it is simple and can achieve significant fraction of the revenue in many cases. A notable such case is when the value distributions for every item are independent. It was shown in [BILW14] that in such a case, this simple mechanism can achieve an $O(\log n)$

fraction of the optimal revenue, i.e. $\text{REV} = O(\log n) \cdot \text{SREV}$. It was also shown there, that $\max\{\text{SREV}, \text{BREV}\}$ can perform even better achieving a constant fraction approximation for the previous case when there is only one bidder. Translating this result to signaling, this immediately implies that in many situations a significant fraction of the optimal revenue can be achieved by the extreme cases of giving full-information to the buyers or no information at all.

Other than these two extreme cases, we can achieve even higher revenue in the Myerson with signaling auction by partitioning items with more general vectors \vec{x}_j . We can partition items into arbitrary disjoint groups and sell each group of items separately, a case studied in the multi-item literature and denoted as PREV. In the most general form, we are allowed to make arbitrary fractional partitions of the items, which we denote by FPREV.

We summarize these connections in the following theorem:

Theorem 11. *The problem of selling a single item whose type is one of n different types occurring with some prior probability $\vec{\pi}$ is equivalent to a multi-item auction setting where there are n different items for sale as in Definition 24. Moreover, different signaling mechanisms are equivalent to particular multi-item auctions:*

- Myerson without Information Revelation *is equivalent to Grand-Bundling auctions that sell all items as one (BREV).*
- Myerson with Full Information Revelation *is equivalent to Selling-Separately (SREV).*
- Myerson with Arbitrary Signaling Schemes *is equivalent to Fractional-Partition mechanisms (FPREV).*
- *The optimal Augmented Auction is equivalent to the Revenue-Optimal multi-item mechanism (REV).*

The following chapter studies how each of these restricted auction forms compare to the optimal achievable revenue by augmented auctions.

Chapter 11

Revenue Comparison

In this chapter, we explore the gaps in revenue achievable in each of the auction formats, i.e. Myerson with no information revelation, Myerson with optimal signaling and the optimal augmented auction. To do this, we exploit the connections to well-studied auction formats in multi-item auction theory presented in Theorem 11.

Using our understanding from Section 10.2, we can immediately get the following relation between the achievable revenue in each of the auction formats:

$$\text{MYERSON} = \text{BREV} \leq \text{PREV} \leq \text{FPREV} = \text{MSIGNALING} \leq \text{REV}$$

We study how large the gap in revenue can get between different auction formats. We examine three different cases that are of particular interest:

- *i.i.d case* - A case where all item types are equally likely and every distribution D_i is product with i.i.d. marginals.
- *independent case* - A case where each D_i is a product distribution with independent (but not necessarily identical) marginals.
- *general case* - A case with arbitrary distributions D_i and $\vec{\pi}$.

| | Myerson (No Information Revelation) vs Myerson with Optimal Signaling | Myerson with Optimal Signaling vs Optimal Augmented Auction |
|-----------------------|--|--|
| i.i.d. | $O(1)$ [Lemma 22] | $O(1)$ [Lemma 22] |
| $\vec{\pi}$ uniform | $\Omega(1)$ [Trivial] | $\Omega(1)$ [Trivial] |
| independent | $O(n)$ [Lemma 23] | $O(1)$ [Lemma 25] |
| $\vec{\pi}$ arbitrary | $\Omega(n)$ [Lemma 24] | $\Omega(1)$ [Trivial] |
| general | $O(n)$ [Lemma 23] | $\leq \infty$ [Trivial] |
| $\vec{\pi}$ arbitrary | $\Omega(n)$ [Lemma 24] | $\geq \infty$ [Lemma 26] |

Table 11.1: Comparison of the worst-case approximation gap among the different auction formats for the case of a **single bidder**.

11.1 Single Bidder

We will show a series of lemmas that bound the worst case revenue gap between different auction formats. The results are summarized in Table 11.1.

Lemma 22. *In a single bidder setting where D_1 is a product distribution with identical marginals and the prior distribution on item types $\vec{\pi}$ is uniform, we have that $\text{REV}(\vec{\pi}) = O(1) \cdot \text{MYERSON}(\vec{\pi})$.*

Proof. Since D_1 is a product distribution with identical marginals and the prior distribution on item types $\vec{\pi}$ is uniform, the corresponding multi-item auction of Definition 24 is a setting with n items and i.i.d. distributions. This case has been studied in [LY13] where it was shown that for a single bidder, bundling can achieve at least a constant fraction of the optimal revenue. This immediately implies that for the optimal augmented auction $\text{REV}(\vec{\pi}) = O(1) \cdot \text{MYERSON}(\vec{\pi})$. \square

Lemma 23. *For any number of bidders, arbitrary distributions D_i and arbitrary prior distribution on item types $\vec{\pi}$, we have that $\text{MSIGNALING}(\vec{\pi}) = O(n) \cdot \text{MYERSON}(\vec{\pi})$.*

Proof. As shown in Lemma 19, the optimal signaling scheme uses at most n signals. Using Equation 10.1, we get:

$$\text{MSIGNALING}(\vec{\pi}) = \max_{\substack{\vec{x}_1 + \dots + \vec{x}_n = \vec{\pi} \\ \vec{x}_j \in \mathbb{R}_{\geq 0}^n}} \sum_{j=1}^n \text{MYERSON}(\vec{x}_j) \leq n \cdot \text{MYERSON}(\vec{\pi})$$

The inequality follows by noting that $\text{MYERSON}(\vec{x}_j) \leq \text{MYERSON}(\vec{\pi})$ since $\vec{x}_j \leq \vec{\pi}$. \square

Lemma 24. *There is a single-bidder instance with a product distribution D_1 , such that $\text{MSIGNALING}(\vec{\pi}) = \Omega(n) \cdot \text{MYERSON}(\vec{\pi})$.*

Proof. To show the lemma, we use a lower bound from [HN12] (Example 15). Consider a uniform prior distribution on item types. Take a large M and let D_1 have support $0, M^i$ with $\Pr(M^i) = M^{-i}$. The revenue achieved by MYERSON (i.e. BREV) is at most $\frac{1}{n}(1 + \frac{1}{M-1})$. In contrast, by always revealing the item type to the buyer, we can sell every item type i at a price M^i getting revenue 1 in expectation. This proves that $\text{MSIGNALING}(\vec{\pi}) = \Omega(n) \cdot \text{MYERSON}(\vec{\pi})$. \square

Lemma 25. *In the case of a single bidder with a product distribution D_1 , we have that $\text{REV}(\vec{\pi}) = O(1) \cdot \text{MSIGNALING}(\vec{\pi})$.*

Proof. As shown in Theorem 11, Myerson with Signaling can achieve revenue at least SREV by full information revelation and BREV by providing no information about the item type. For a single additive bidder with independent distributions, it is shown in [BILW14], that $\text{REV} \leq 6 \cdot \max\{\text{SREV}, \text{BREV}\}$. This immediately implies that $\text{REV}(\vec{\pi}) = O(1) \cdot \text{MSIGNALING}(\vec{\pi})$. \square

Lemma 26. *For any $a \geq 1$, there exists a single-bidder instance with $n = 2$ item types such that:*

$$\text{REV}(\vec{\pi}) \geq a \cdot \text{MSIGNALING}(\vec{\pi})$$

Proof. In Lemma 23, we showed that $\text{MSIGNALING}(\vec{\pi}) = O(n) \cdot \text{MYERSON}(\vec{\pi})$. Since $\text{MYERSON} = \text{BREV}$, we get that for two items $\text{MSIGNALING} = O(1) \cdot \text{BREV}$. The lemma follows since BREV and REV can have an arbitrary gap in revenue even for two items as shown in [HN13]. \square

11.2 Multiple Bidders

Working similarly for multiple bidders as in the case of one bidder, we show a series of lemmas that bound the worst case revenue gap between different auction formats.

| | Myerson (No Information Revelation) vs Myerson with Optimal Signaling | Myerson with Optimal Signaling vs Optimal Augmented Auction |
|-----------------------|--|--|
| i.i.d. | $O(n)$ [Lemma 23] | $O(\log n)$ [Lemma 28] |
| $\vec{\pi}$ uniform | $\Omega(\frac{n}{\log n})$ [Lemma 27] | $\Omega(\frac{\log n}{\log \log n})$ [Lemma 29] |
| independent | $O(n)$ [Lemma 23] | $O(\log n)$ [Lemma 28] |
| $\vec{\pi}$ arbitrary | $\Omega(n)$ [Lemma 24] | $\Omega(\frac{\log n}{\log \log n})$ [Lemma 29] |
| general | $O(n)$ [Lemma 23] | $\leq \infty$ [Trivial] |
| $\vec{\pi}$ arbitrary | $\Omega(n)$ [Lemma 24] | $\geq \infty$ [Lemma 26] |

Table 11.2: Comparison of the worst-case approximation gap among the different auction formats for the case of **many bidders**.

The results are summarized in Table 11.2.

Lemma 27. *There exists an instance where every D_i is a product distribution with identical marginals and the prior distribution on item types $\vec{\pi}$ is uniform, such that $\text{MSIGNALING}(\vec{\pi}) = \Omega(\frac{n}{\log n}) \cdot \text{MYERSON}(\vec{\pi})$.*

Proof. Consider an instance with $m = n$ bidders such that each bidder has value 1 for item type j with probability $\frac{1}{n}$ and 0 otherwise. The distribution over item types is uniform. By a simple tail bound for the Binomial distribution, we can see that a single bidder has non-zero value for $\Omega(\log n)$ items types with probability at most $\frac{1}{n^2}$. Taking a union bound over all bidders, we get that with probability $1 - \frac{1}{n}$, every bidder has at most $O(\log n)$ non-zero values and thus the maximum revenue achievable in this case by MYERSON is $O(\frac{\log n}{n})$ (recall that each type appears with probability $\frac{1}{n}$). In contrast, if the seller always reveals the type of the item, the achieved revenue is $O(1)$, since he can sell the item at price 1 which is bought with constant probability $\approx 1 - \frac{1}{e}$. \square

Lemma 28. *In the case where all distributions D_i are identical product distributions, we have that $\text{REV}(\vec{\pi}) = O(\log n) \cdot \text{MSIGNALING}(\vec{\pi})$.*

Proof. In this setting (independent case), [BILW14] show that $\text{REV} \leq (\ln n + 6) \cdot \text{SREV}$. The lemma immediately follows as for the Myerson with Signaling auction, we have $\text{MSIGNALING} \geq \text{SREV}$. \square

Lemma 29. *There exists an instance where every D_i is a product distribution with identical marginals and the prior distribution on item types $\vec{\pi}$ is uniform, such that $\text{REV}(\vec{\pi}) = \Omega\left(\frac{\log n}{\log \log n}\right) \cdot \text{MSIGNALING}(\vec{\pi})$.*

Proof. To show the lower bound on optimal revenue, we will construct an example where the optimal multi-item auction gets revenue $\text{REV} = \Omega(n \log n)$, while the optimal fractional partition mechanism FPREV that corresponds to MSIGNALING , as shown in Theorem 11, gets revenue $\text{FPREV} = O(n \log \log n)$.

The example is the same as in [BILW14] (Proposition 8), where it was used to show that $\text{REV} = \Omega(\log n) \text{PREV}$. There are n items and \sqrt{n} bidders and the value of each bidder for each item is sampled with probability $\frac{1}{\sqrt{n}}$ from a distribution F with CDF¹

$$F(x) = \begin{cases} 1 - x^{-1} & \text{if } x \in [1, n^{1/8}] \\ 1 & \text{if } x > n^{1/8} \end{cases}$$

and with the remaining probability it is 0. In [BILW14], it is shown that $\text{REV} = \Omega(n \log n)$. We will show that $\text{FPREV} = O(n \log \log n)$ by showing that for any fractional bundle \vec{f} that contains f_j units of item j , selling the bundle gives revenue $\text{BREV}(\vec{f}) = O(\log \log n) \cdot \sum_{j=1}^n f_j$.

Without loss of generality, assume that $\max_j f_j = 1$. Now consider the fractional bundle \vec{f}' by rounding up all values of \vec{f} to the smallest value 2^{-a} for a in the set $\{0, 1, \dots, \log n\}$. It is clear that the revenue from the fractional bundle \vec{f}' is at least as much as the revenue from the fractional bundle \vec{f} since all fractional values increased. Moreover, we have that $\sum_j f'_j = \sum_{j:f_j > n^{-1}} f'_j + \sum_{j:f_j \leq n^{-1}} f'_j \leq \sum_{j:f_j > n^{-1}} 2f_j + 1$. So $\sum_j f'_j = O(1) \cdot \sum_j f_j$.

We now bound the revenue of the fractional bundle \vec{f}' . The vector \vec{f}' consists of at most $\log n$ different values. For every such value $a \in \{0, 1, \dots, \log n\}$, we let S_a be the set of all items j with $f'_j = 2^{-a}$. We now consider a different setting with $\log n$ items, where item a corresponds to 2^{-a} units of the items in S_a . In this setting, there are $\log n$ items and we know that selling the items separately is a $O(\log \log n)$ approximation

¹ F is an Equal-Revenue distribution with all mass above $n^{1/8}$ moved to an atom at $n^{1/8}$.

to the optimal revenue and in particular a $O(\log \log n)$ approximation to selling the grand bundle which corresponds to the revenue achieved by the fractional bundle \vec{f} in the original setting.

This implies that $\text{BREV}(\vec{f}') \leq O(\log \log n) \sum_{a=0}^{\log n} 2^{-a} \text{BREV}(S_a)$ where $\text{BREV}(S_a)$ is the revenue achievable by selling the bundle of items in S_a . As shown in [BILW14] (Proposition 8), $\text{BREV}(S_a) \leq O(1)|S_a|$ which implies that $\text{BREV}(\vec{f}) \leq \text{BREV}(\vec{f}') \leq O(\log \log n)|\vec{f}'| = O(\log \log n)|\vec{f}|$ as required. \square

Chapter 12

Discussion and Extensions

Our work studies the problem of signaling through the lens of multidimensional mechanism design in a model of valuations inspired by ad auctions. We show that the optimal co-design of signaling and auction maps to an optimal multi-item auction design in an associated multi-item setting with additive bidders, as specified in Definition 24. Our reduction allows us to leverage our understanding of multi-item auctions to (i) design optimal augmented auctions; (ii) understand how stratagems such as targeting, discount and insurance can be used to increase revenue; and (iii) quantify the suboptimality of signaling schemes studied in the literature by relating them to suboptimal multi-item auction designs in the associated multi-item setting. In this chapter, we discuss additional benefits from our reduction, relating to multi-item extensions, computation and structure.

12.1 Multiple-Items

Since our mapping to multi-item auctions is already multi-dimensional, we can easily accommodate multiple items in our framework for augmented auction design as follows: in a setting with m bidders, k items with n types each, the optimal augmented auction design reduces to a multi-item auction design problem with m additive bidders and $k \cdot n$ items. (The proof of this is identical to that of Theorem 10.) Multi-item settings are important in applications of online ads where there are several ad-slots

on a webpage and advertisers may have different values for each of these slots, which may also depend on the type of user that visits the page. For example, an advertiser can have values $(5, 3)$ for the slots on the top and the bottom of the page if the user is under 25 years old and values $(8, 2)$ otherwise.

12.2 Structure and Computation

By reducing the augmented auction design problem to that of revenue-optimal multi-item auctions, we get immediate implications about the structure of the problem and its computational complexity.

We can use the tools and techniques developed in Part 1 for the single-bidder case to immediately get structural characterizations for the optimal single-bidder augmented auction. Our characterization for grand-bundling optimality gives necessary and sufficient conditions for the Myerson with No Information Revelation mechanism to be optimal while optimality of more general mechanisms can be checked by verifying that the appropriate stochastic dominance conditions are satisfied.

In terms of computation, the results of [AFH⁺12, CDW12a, CDW12b] imply that finding the optimal BIC augmented auction can be done in time polynomial in the support size of all bidder distributions D_i . At the same time, when all distributions D_i are product distributions and are given implicitly by listing their marginals (as in the i.i.d. and independent settings discussed in Chapter 11) the support size becomes exponential in the number of item types, and we can leverage our reduction to the multi-item case and the computational complexity results of Chapter 7 to show that the problem is #P-hard. Similarly, our connection to the multi-item case (in particular Theorem 11) and the results of [DDT12] imply that Myerson with No Information Revelation is also #P-hard. Still, when all D_i are product distributions, we can leverage recent work of Babaioff et al [BILW14] and Yao [Yao15] for approximately optimal multi-item auctions to obtain an augmented auction achieving a constant fraction of the optimal revenue.

12.3 Correlation Among Bidders and Item Types

So far, we have assumed that the bidder values are independent and that the prior on the item types is the same for all bidders. Following the work of Milgrom and Weber, the more general model presented in Section 9.1 does not make any independence assumptions and allows for arbitrary correlation between item and bidder types. Our reduction can be easily extended to capture such situations by reducing the problem again to a multi-item auction where bidders have interdependent values.

Definition 25 (Optimal Multi-Item Auction with Interdependent Values). *We have m additive bidders and n items. Every bidder has a type t_i . Bidder i 's value for item j depends both on his type and the types of other bidders. The type vector of $\vec{t} = (t_1, \dots, t_m)$ comes from a joint distribution D .*

Similar to Theorem 10, this setting is equivalent to a full-information multi-item auction setting with additive bidders as in Definition 25. A bidder i with type t_i in the original setting has value $V_i(j, \vec{t}) \cdot \text{Prob}_{\tau_{item} \sim D|\vec{t}}[\tau_{item} = j]$ for every item j in the multi-item auction when all other bidders have values \vec{t}_{-i} . Notice that for the case of a single bidder, the setting is still identical to the multi-item setting with private values that we considered in the first part of this work. When multiple bidders are involved the equivalent multi-item auction design problem is for a setting where bidders have interdependent values. Such problems are usually much harder to solve in general as even for single item auctions Myerson's theory does not apply and optimal auctions can have very complex structure.

We now show this more general equivalence between augmented auctions in two-sided information asymmetry settings and multi-item auctions where bidders have interdependent values. The proof of this more general equivalence uses the revelation principle and is identical to the proof of Theorem 10. We repeat it here with the appropriate changes for completeness.

Proof. Consider a mechanism \mathcal{M} that is ex-post implementable. The mechanism \mathcal{M} may use multiple rounds of communication and information revelation to the buyers.

For each $i \in [m]$, let $A_i(t_i)$ be the (possibly randomized) equilibrium strategy of buyer i when his type is t_i . Each strategy $A_i(t_i)$ specifies the actions that bidder i takes in the mechanism \mathcal{M} given information about the item type and the actions of other buyers, in order to maximize buyer i 's expected payoff.

Let $\mathcal{P}_i(j, \vec{A})$ be an indicator random variable that indicates whether buyer i gets the item when buyers choose strategies $\vec{A} = (A_1, \dots, A_m)$ and the realized item type is $j \in [n]$. Similarly, let $\mathcal{T}_i(j, \vec{A})$ denote the price buyer i is asked to pay. The expected utility of buyer i is then equal to

$$\mathbb{E}_{j \sim D|\vec{t}}[\mathbb{E}[\mathcal{P}_i(j, \vec{A})V_i(j, \vec{t}) - \mathcal{T}_i(j, \vec{A})]]$$

where the first expectation is with respect to the randomness of the item type according to bidder i , while the second expectation is with respect to the randomness in the choices of the mechanism, the information revealed and the actions of the buyers under the mixed strategies \vec{A} .

We now examine the equilibrium conditions for the buyer strategies. For all possible types $\vec{t} = (t_1, \dots, t_m)$ and all possible misreports t'_i for buyer i , we let $\vec{A}(\vec{t}) = (A_1(t_1), \dots, A_m(t_m))$ and $\vec{A}(t'_i, \vec{t}_{-i}) = (A_1(\vec{v}_1), \dots, A_i(\vec{v}'_i), \dots, A_m(\vec{v}_m))$. The equilibrium conditions for the strategy profile $\vec{A}(\vec{t})$ imply that:

$$\begin{aligned} \mathbb{E}_{j \sim D|\vec{t}}[\mathbb{E}[\mathcal{P}_i(j, \vec{A}(\vec{t}))V_i(j, \vec{t}) - \mathcal{T}_i(j, \vec{A}(\vec{t}))]] &\geq \\ \mathbb{E}_{j \sim D|\vec{t}}[\mathbb{E}[\mathcal{P}_i(j, \vec{A}(t'_i, \vec{t}_{-i}))V_i(j, \vec{t}) - \mathcal{T}_i(j, \vec{A}(t'_i, \vec{t}_{-i}))]] & \end{aligned}$$

Now set the variables $x_{ij}(\vec{t}) = \mathbb{E}[\mathcal{P}_i(j, \vec{A}(\vec{t}))]$ and $c_i(\vec{t}) = \mathbb{E}_{j \sim D|\vec{t}}[\mathbb{E}[\mathcal{T}_i(j, \vec{A}(\vec{t}))]]$. The equation above can be rewritten for all i, \vec{t} and potential misreports t'_i as:

$$\begin{aligned} \sum_{j=1}^n x_{ij}(\vec{t})V_i(j, \vec{t}) \cdot \text{Prob}_{\tau_{item} \sim D|\vec{t}}[\tau_{item} = j] - c_i(\vec{t}) &\geq \tag{IC} \\ \sum_{j=1}^n x_{ij}(t'_i, \vec{t}_{-i})V_i(j, \vec{t}) \cdot \text{Prob}_{\tau_{item} \sim D|\vec{t}}[\tau_{item} = j] - c_i(t'_i, \vec{t}_{-i}) & \end{aligned}$$

Moreover, the mechanism must satisfy the individual rationality constraint in order

for the bidders to participate

$$\sum_{j=1}^n x_{ij}(\vec{t}) V_i(j, \vec{t}) \cdot \text{Prob}_{\tau_{item} \sim D | \vec{t}}[\tau_{item} = j] - c_i(\vec{t}) \geq 0 \quad (\text{IR})$$

and that for all $j \in [n]$: $\sum_{i=1}^m x_{ij}(\vec{t}) \leq 1$. A mechanism that satisfies all these constraints and achieves expected revenue

$$\mathbb{E}_{(j, \vec{t}) \sim D} \left[\sum_{i=1}^m c_i(\vec{t}) \right]$$

can be viewed as a multi-item direct auction for m additive bidders with n items where the value of a bidder with type t_i for item j is simply $V_i(j, \vec{t}) \cdot \text{Prob}_{\tau_{item} \sim D | \vec{t}}[\tau_{item} = j]$. This direct mechanism allocates the item j to bidder i with probability $x_{ij}(\vec{t})$ when all bidders have types \vec{t} . The **(IC)** constraint implies that the auction is incentive compatible while the **(IR)** constraint ensures bidder participation. Any direct mechanism \mathcal{M} for the multi-item setting is a feasible augmented auction since the seller can use \mathcal{M} to presell all item types before observing the real item type and then decide on the final allocation after the type is observed.

The same reduction goes through for converting any Bayes-Nash equilibrium to an equivalent Bayesian incentive-compatible direct mechanism that presells all the items. The difference is that in the **(IC)** and **(IR)** constraints above for bidder i , we take an expectation over all bidder types \vec{t}_{-i} other than i . \square

An intermediate setting is when the item type is independent on the bidder types and each bidder's value depends only on his own type and the item type. In such cases, the equivalent multi-item auction corresponds to a private value setting with correlated values among bidder.

Chapter 13

Conclusions and Future Directions

In the past few decades, there has been remarkable progress in our understanding of auction theory. Building on Myerson's result [Mye81], virtually every aspect of revenue optimization in single-item settings has been explored.

Unfortunately, multi-item revenue optimization has proven significantly more challenging and did not enjoy the same fate. Even though there is an increasing number of applications that are inherently multi-item, our understanding of optimal mechanisms is still very limited. Indeed multi-item mechanisms exhibit a very rich structure and often display very unintuitive properties.

In this work, we were able to connect the problem of multi-item revenue optimization for a single buyer to a well-studied problem in mathematics, designing optimal transportation maps. Through this connection, based on duality theory and optimization, we obtained a fresh perspective on multi-item mechanism design. We were able to characterize the structure of multi-item mechanisms and develop an analytical framework via which the optimal mechanism can be identified.

In addition to our work in multi-item mechanisms, in the second part, we explored settings where bidders have uncertainty about the item for sale. We uncovered a strong connection between this setting and the optimal multi-item mechanism design problem in the first part. Through this connection, we were able to readily export our tools for finding optimal mechanisms. Moreover, we were able to quantify the revenue loss depending on what information the auctioneer reveals about the type of

the item prior to running the auction.

We believe that our results are opening exciting and promising directions for future investigation. To identify just a few:

- It is important to extend the duality-based framework we developed in Part 1 to accommodate multiple bidders. Optimal multi-bidder multi-item mechanism design problems are unsolved even for very simple examples. One such example is a setting with 2 items and 2 additive bidders where each bidder's value for each item is drawn independently from $\text{Uniform}[0, 1]$.
- What is the sensitivity of the structural results on the details of the bidder type distributions? As the structure of optimal mechanisms might change significantly with small changes in the specification, a possible approach for robust mechanism design is to ask for mechanisms that maximize expected revenue for the worst distribution from some class. A related result by Gabriel Carroll [Car17] considers the class of all joint distributions with given marginals.
- In settings with bidder uncertainty, when can the optimal mechanism be implemented using insurance policies? The equivalence of this setting to multi-item auctions required that types are observable by the auctioneer. However, in several examples there exist implementations through insurance in which the auctioneer never observes the true type of the item.

Bibliography

- [AABG13] Ittai Abraham, Susan Athey, Moshe Babaioff, and Michael Grubb. Peaches, lemons, and cookies: designing auction markets with dispersed information. In *Proceedings of the 14th ACM conference on Electronic commerce*, pages 7–8. ACM, 2013. [8](#)
- [AFH⁺12] Saeed Alaei, Hu Fu, Nima Haghpanah, Jason Hartline, and Azarakhsh Malekian. Bayesian Optimal Auctions via Multi- to Single-agent Reduction. In *the 13th ACM Conference on Electronic Commerce (EC)*, 2012. [2.2](#), [7](#), [12.2](#)
- [Ake70] George A Akerlof. The market for “lemons”: Quality uncertainty and the market mechanism. *The quarterly journal of economics*, pages 488–500, 1970. [7](#)
- [AN87] Edward J Anderson and Peter Nash. *Linear programming in infinite-dimensional spaces: theory and applications*. John Wiley & Sons, 1987. [2.1.1](#)
- [Arm96] Mark Armstrong. Multiproduct nonlinear pricing. *Econometrica: Journal of the Econometric Society*, pages 51–75, 1996. [2.2](#)
- [Arm99] Mark Armstrong. Price discrimination by a many-product firm. *The Review of Economic Studies*, 66(1):151–168, 1999. [2.2](#)
- [BB99] Yannis Bakos and Erik Brynjolfsson. Bundling information goods: Pricing, profits, and efficiency. *Management Science*, 45(12):1613–1630, 1999. [1.1](#), [2](#)
- [BCKW10] P Briest, S Chawla, R Kleinberg, and S M Weinberg. Pricing randomized allocations. In *the 21st Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2010. [2.2](#)
- [BFL91] László Babai, Lance Fortnow, and Carsten Lund. Non-Deterministic Exponential Time has Two-Prover Interactive Protocols. *Computational Complexity*, 1:3–40, 1991. [7.6.2](#)
- [BHV14] Kshipra Bhawalkar, Patrick Hummel, and Sergei Vassilvitskii. Value of targeting. In *Algorithmic Game Theory*, pages 194–205. Springer, 2014. [8](#)

- [BILW14] M Babaioff, N Immorlica, B Lucier, and S M Weinberg. A simple and approximately optimal mechanism for an additive buyer. In *the 55th Annual Symposium on Foundations of Computer Science (FOCS)*, 2014. [2.2](#), [10.2](#), [11.1](#), [11.2](#), [11.2](#), [12.2](#)
- [BMS12] Peter Bro Miltersen and Or Sheffet. Send mixed signals: earn more, work less. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, pages 234–247. ACM, 2012. [8](#), [9.2](#)
- [Boa09] Simon Board. Revealing information in auctions: the allocation effect. *Economic Theory*, 38(1):125–135, 2009. [8](#)
- [BP07] Dirk Bergemann and Martin Pesendorfer. Information structures in optimal auctions. *Journal of Economic Theory*, 137(1):580–609, 2007. [8](#)
- [Bri08] Patrick Briest. Uniform budgets and the envy-free pricing problem. In *the 35th International Colloquium in Automata, Languages and Programming (ICALP)*, 2008. [2.2](#)
- [BW15] Dirk Bergemann and Achim Wambach. Sequential information disclosure in auctions. *Journal of Economic Theory*, 159:1074–1095, 2015. [8](#)
- [Car17] Gabriel Carroll. Robustness and separation in multidimensional screening. *Econometrica*, 85(2):453–488, 2017. [13](#)
- [CCD⁺15] Yu Cheng, Ho Yee Cheung, Shaddin Dughmi, Ehsan Emamjomeh-Zadeh, Li Han, and Shang-Hua Teng. Mixture selection, mechanism design, and signaling. In *the 56th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2015. [8](#)
- [CDO⁺15] Xi Chen, Ilias Diakonikolas, Anthi Orfanou, Dimitris Paparas, Xiaorui Sun, and Mihalis Yannakakis. On the complexity of optimal lottery pricing and randomized mechanisms. In *Foundations of Computer Science (FOCS), 2015 IEEE 56th Annual Symposium on*, pages 1464–1479. IEEE, 2015. [2.2](#)
- [CDP⁺14] Xi Chen, Ilias Diakonikolas, Dimitris Paparas, Xiaorui Sun, and Mihalis Yannakakis. The complexity of optimal multidimensional pricing. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1319–1328. Society for Industrial and Applied Mathematics, 2014. [2.2](#)
- [CDW12a] Yang Cai, Constantinos Daskalakis, and S. Matthew Weinberg. An Algorithmic Characterization of Multi-Dimensional Mechanisms. In *the 44th Annual ACM Symposium on Theory of Computing (STOC)*, 2012. [2.2](#), [7](#), [12.2](#)

- [CDW12b] Yang Cai, Constantinos Daskalakis, and S. Matthew Weinberg. Optimal Multi-Dimensional Mechanism Design: Reducing Revenue to Welfare Maximization. In *the 53rd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2012. [2.2](#), [7](#), [12.2](#)
- [CDW13] Yang Cai, Constantinos Daskalakis, and S. Matthew Weinberg. Reducing Revenue to Welfare Maximization : Approximation Algorithms and other Generalizations. In *the 24th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2013. [2.2](#)
- [CH13] Yang Cai and Zhiyi Huang. Simple and nearly optimal multi-item auctions. In *the 24th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2013. [2.2](#)
- [DDT12] Constantinos Daskalakis, Alan Deckelbaum, and Christos Tzamos. Optimal pricing is hard. In *the 8th Workshop on Internet & Network Economics (WINE)*, 2012. [2.2](#), [12.2](#)
- [DDT13] Constantinos Daskalakis, Alan Deckelbaum, and Christos Tzamos. Mechanism design via optimal transport. In *14th ACM Conference on Electronic Commerce (EC)*, 2013. [1.1](#), [1.3](#), [2.1](#), [2.1.1](#), [2.1.2](#), [2.2](#), [4.1](#), [2](#), [4.3.2](#), [5.1](#), [6.4](#)
- [DDT14] Constantinos Daskalakis, Alan Deckelbaum, and Christos Tzamos. The complexity of optimal mechanism design. In *25th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2014. [1.3](#)
- [DDT17] Constantinos Daskalakis, Alan Deckelbaum, and Christos Tzamos. Strong Duality for a Multiple-Good Monopolist. *Econometrica (accepted)*, 2017. [1.3](#)
- [DFK11] Shahar Dobzinski, Hu Fu, and Robert D. Kleinberg. Optimal Auctions with Correlated Bidders are Easy. In *the 43rd ACM Symposium on Theory of Computing (STOC)*, 2011. [2.2](#), [7.6.1](#), [7.6.1](#)
- [DIR14] Shaddin Dughmi, Nicole Immorlica, and Aaron Roth. Constrained signaling in auction design. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1341–1357. SIAM, 2014. [8](#)
- [DPT16] Constantinos Daskalakis, Christos Papadimitriou, and Christos Tzamos. Does information revelation improve revenue? In *Proceedings of the 2016 ACM Conference on Economics and Computation*, pages 233–250. ACM, 2016. [1.3](#)
- [Dud02] R.M. Dudley. *Real Analysis and Probability*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2002. [5.3.1](#), [5.3.1](#)

- [EFG⁺14] Yuval Emek, Michal Feldman, Iftah Gamzu, Renato PaesLeme, and Moshe Tennenholtz. Signaling schemes for revenue maximization. *ACM Transactions on Economics and Computation*, 2(2):5, 2014. [8](#), [9.2](#)
- [Eke10] Ivar Ekeland. Notes on optimal transportation. *Economic Theory*, 42(2):437–459, 2010. [2.2](#)
- [ES07] Péter Esö and Balazs Szentes. Optimal information disclosure in auctions and the handicap auction. *The Review of Economic Studies*, 74(3):705–731, 2007. [8](#)
- [FJM⁺12] Hu Fu, Patrick Jordan, Mohammad Mahdian, Uri Nadav, Inbal Talgam-Cohen, and Sergei Vassilvitskii. Ad auctions with data. In *Symposium on Algorithmic Game Theory (SAGT)*, 2012. [8](#), [6](#), [9.2](#), [9.4](#), [10.1.3](#)
- [FKM11] Alessio Figalli, Young-Heon Kim, and Robert J McCann. When is multi-dimensional screening a convex program? *Journal of Economic Theory*, 146(2):454–478, 2011. [2.2](#)
- [GK14] Yiannis Giannakopoulos and Elias Koutsoupias. Duality and optimality of auctions for uniform distributions. In *the 15th ACM Conference on Electronic Commerce (EC)*, 2014. [1.1](#), [2.1](#), [2.1.1](#), [2.2](#), [4.1](#), [2](#), [4.3.2](#), [5.1](#)
- [GMMV15] Arpita Ghosh, Mohammad Mahdian, R Preston McAfee, and Sergei Vassilvitskii. To match or not to match: Economics of cookie matching in online advertising. *ACM Transactions on Economics and Computation*, 3(2):12, 2015. [8](#)
- [HH15] Nima Haghpanah and Jason Hartline. Reverse mechanism design. In *Proceedings of the 16th ACM Conference on Economics and Computation*, pages 757–758. ACM, 2015. [2.2](#)
- [HM15] Patrick Hummel and R Preston McAfee. When does improved targeting increase revenue? In *Proceedings of the 24th International Conference on World Wide Web*, pages 462–472. International World Wide Web Conferences Steering Committee, 2015. [8](#)
- [HN12] Sergiu Hart and Noam Nisan. Approximate Revenue Maximization with Multiple Items. In *the 13th ACM Conference on Electronic Commerce (EC)*, 2012. [2](#), [2.2](#), [11.1](#)
- [HN13] Sergiu Hart and Noam Nisan. The menu-size complexity of auctions. In *14th ACM Conference on Electronic Commerce (EC)*, 2013. [2.2](#), [6.3](#), [11.1](#)
- [HR15] Sergiu Hart and Philip J Reny. Maximal revenue with multiple goods: Nonmonotonicity and other observations. *Theoretical Economics*, 10(3):893–922, 2015. [2.2](#)

- [Lue68] David G Luenberger. *Optimization by vector space methods*. John Wiley & Sons, 1968. [2.1.1](#)
- [LY13] Xinye Li and Andrew Chi-Chih Yao. On revenue maximization for selling multiple independently distributed items. *Proceedings of the National Academy of Sciences*, 110(28):11232–11237, 2013. [2.2](#), [11.1](#)
- [MHJ15] Domenico Menicucci, Sjaak Hurkens, and Doh-Shin Jeon. On the optimality of pure bundling for a monopolist. *Journal of Mathematical Economics*, 60:33–42, 2015. [2.2](#)
- [MMW89] R. Preston McAfee, John McMillan, and Michael D. Whinston. Multi-product monopoly, commodity bundling, and correlation of values. *The Quarterly Journal of Economics*, pages 371–383, 1989. [1.1](#), [2](#)
- [MR84] Eric Maskin and John Riley. Monopoly with incomplete information. *The RAND Journal of Economics*, 15(2):171–196, 1984. [2](#)
- [MV06] A.M. Manelli and D.R. Vincent. Bundling as an optimal selling mechanism for a multiple-good monopolist. *Journal of Economic Theory*, 127(1):1–35, 2006. [1.1](#), [2](#), [2.1](#), [2.1.1](#), [2.1.2](#), [2.2](#), [3.1.2](#), [3.1.2](#), [4.1](#), [4.3.1](#), [5.1](#)
- [MV07] A. M. Manelli and D. R. Vincent. Multidimensional Mechanism Design: Revenue Maximization and the Multiple-Good Monopoly. *Journal of Economic Theory*, 137(1):153–185, 2007. [2.2](#)
- [MW82] Paul R Milgrom and Robert J Weber. A theory of auctions and competitive bidding. *Econometrica*, pages 1089–1122, 1982. [8](#), [9.1](#)
- [Mye79] Roger B Myerson. Incentive compatibility and the bargaining problem. *Econometrica: journal of the Econometric Society*, pages 61–73, 1979. [3.1.1](#), [7.2](#)
- [Mye81] Roger B. Myerson. Optimal Auction Design. *Mathematics of Operations Research*, 6(1):58–73, 1981. [1](#), [1](#), [1.2](#), [2](#), [4.2](#), [13](#)
- [NRTV07] Noam Nisan, Tim Roughgarden, E. Tardos, and V. V. Vazirani, editors. *Algorithmic Game Theory*. Cambridge University Press, 2007. [6](#), [7.1](#), [7.2](#)
- [Pav11] Gregory Pavlov. Optimal mechanism for selling two goods. *The B.E. Journal of Theoretical Economics*, 11(1):1–35, February 2011. [2.1.1](#), [2.1.2](#), [4.3.2](#), [3](#)
- [PP11] Christos H. Papadimitriou and George Pierrakos. On optimal single-item auctions. In *the 43rd ACM Symposium on Theory of Computing (STOC)*, 2011. [2.2](#)
- [RC98] Jean-Charles Rochet and Philippe Choné. Ironing, sweeping, and multidimensional screening. *Econometrica*, 66(4):783–826, July 1998. [2.2](#), [3.1.2](#), [4.1](#), [4.1.2](#)

- [Roc87] Jean-Charles Rochet. A necessary and sufficient condition for rationalizability in a quasi-linear context. *Journal of Mathematical Economics*, 16(2):191–200, April 1987. [2.1.1](#), [2.2](#), [3.1.2](#)
- [RS81] John G Riley and William F Samuelson. Optimal auctions. *The American Economic Review*, 71(3):381–392, 1981. [2](#)
- [RS03] Jean-Charles Rochet and Lars A Stole. The economics of multidimensional screening. *Econometric Society Monographs*, 35:150–197, 2003. [2.2](#)
- [RZ83] John Riley and Richard Zeckhauser. Optimal selling strategies: When to haggle, when to hold firm. *The Quarterly Journal of Economics*, pages 267–289, 1983. [2](#)
- [Sha92] Adi Shamir. $IP = PSPACE$. *J. ACM*, 39(4):869–877, October 1992. [7.6.2](#)
- [Sip06] Michael Sipser. *Introduction to the Theory of Computation: Second Edition*. PWS Pub., Boston, 2 edition, 2006. [7.5](#)
- [SS10] M. Shaked and J.G. Shanthikumar. *Stochastic Orders*. Springer Series in Statistics. Springer, 2010. [5.3.1](#)
- [Tha04] J. Thanassoulis. Hagglng over substitutes. *Journal of Economic Theory*, 117(2):217–245, 2004. [1.1](#), [2](#), [2.2](#)
- [Tza] Christos Tzamos. Optimal Mechanisms for Beta Distributions. <http://christos.me/betas>. [6.3](#)
- [Vil08] Cédric Villani. *Optimal Transport: Old and New*, volume 338. Springer Science & Business Media, 2008. [2.1.1](#), [2.2](#), [4.1.2](#), [4.4.1](#)
- [Yao15] Andrew Chi-Chih Yao. An n -to-1 Bidder Reduction for Multi-item Auctions and its Applications. In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 92–109. SIAM, 2015. [2.2](#), [12.2](#)