Strong Duality for a Multiple-Good Monopolist

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1145/2764468.2764539">http://dx.doi.org/10.1145/2764468.2764539</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>Association for Computing Machinery (ACM)</td>
</tr>
<tr>
<td>Version</td>
<td>Original manuscript</td>
</tr>
<tr>
<td>Accessed</td>
<td>Wed Dec 19 04:00:48 EST 2018</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/110829">http://hdl.handle.net/1721.1/110829</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Creative Commons Attribution-Noncommercial-Share Alike</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td><a href="http://creativecommons.org/licenses/by-nc-sa/4.0/">http://creativecommons.org/licenses/by-nc-sa/4.0/</a></td>
</tr>
</tbody>
</table>
Strong Duality for a Multiple-Good Monopolist

Constantinos Daskalakis∗ Alan Deckelbaum† Christos Tzamos‡
EECS, MIT Math, MIT EECS, MIT

April 29, 2016

Abstract

We provide a duality-based framework for revenue maximization in a multiple-good monopoly. Our framework shows that every optimal mechanism has a certificate of optimality, taking the form of an optimal transportation map between measures. Using our framework, we prove that grand-bundling mechanisms are optimal if and only if two stochastic dominance conditions hold between specific measures induced by the buyer’s type distribution. This result strengthens several results in the literature, where only sufficient conditions for grand-bundling optimality have been provided. As a corollary of our tight characterization of grand-bundling optimality, we show that the optimal mechanism for \( n \) independent uniform items each supported on \([c, c + 1]\) is a grand-bundling mechanism, as long as \( c \) is sufficiently large, extending Pavlov’s result for 2 items [Pav11]. In contrast, our characterization also implies that, for all \( c \) and for all sufficiently large \( n \), the optimal mechanism for \( n \) independent uniform items supported on \([c, c + 1]\) is not a grand bundling mechanism. 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.

Keywords: Revenue maximization, mechanism design, strong duality, grand bundling

∗Supported by a Sloan Foundation Fellowship, a Microsoft Research Faculty Fellowship, and NSF Awards CCF-0953960 (CAREER) and CCF-1101491.
†Supported by Fannie and John Hertz Foundation Daniel Stroock Fellowship and NSF Award CCF-1101491.
‡Supported by NSF Award CCF-1101491.
1 Introduction

We study the problem of revenue maximization for a multiple-good monopolist. Given $n$ heterogeneous goods and a probability distribution $f$ over $\mathbb{R}^n_{\geq 0}$, 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 [Mye81], 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 $f$. For general $n$, it has been known that the optimal mechanism may exhibit much richer structure. 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 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. See, e.g., [RS03] for a comprehensive survey of work spanning our problem, as well as [MV07] and [FKM11] for additional references.

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. 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 guaranteed to apply to arbitrary settings of $n$ and $f$ (with mild assumptions such as differentiability). In particular, we strengthen prior work [DDT13, MV06, GK14], which has identified optimal multi-item mechanisms only in special cases. Importantly, we can leverage our strong duality to characterize optimal multi-item mechanisms. From a technical standpoint we provide new analytical methodology for multi-dimensional mechanism design when strict convexity is lacking. When strict convexity is present due to, say, strictly convex production costs, first order conditions suffice to characterize optimal mechanisms [RC98]. We proceed to discuss our contributions in detail, providing a roadmap to the paper.

Strong Duality. Our main result (presented in Theorem 2) formulates the dual of the optimal mechanism design problem and establishes strong duality between the two problems (i.e. that the optimal values of the two optimization problems are identical). Our approach is as follows:

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 her 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 $\mu$ over the type space of the buyer, obtained by an application of the divergence theorem (whose use is standard in this context). Measure $\mu$
is easily defined in terms of the buyer’s type distribution $f$ (see Equation (1)). Our formulation is summarized in Theorem 1, while Section 2.2 illustrates our formulation in the basic setting of independent uniform items.

In Theorem 2, we formulate a dual in the form of an optimal transport problem, and establish strong duality between the two problems. Roughly speaking, our dual formulation is given the signed measure $\mu$ (from Theorem 1) and solves the following minimization problem: (i) first, it is allowed to massage $\mu$ into a measure $\mu'$ that dominates $\mu$ with respect to a particular “convex dominance” stochastic order; (ii) second, it is supposed to find a coupling of the positive part $\mu'_+$ of $\mu'$ with its negative part $\mu'_-$; (iii) if a unit of mass of $\mu'_+$ at $x$ is coupled with a unit of mass of $\mu'_-$ at $y$, we are charged $\|x - y\|_1$. The goal is to minimize the expected 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 [Vil09], 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. The proof is presented in Section A, but it is not necessary to understand the other results in this paper. 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.

Using our framework, we can provide shorter proofs of optimality of known mechanisms. As an illustrating example, we show, in Section 5.1, 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 5.2, 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 (c.f. previous work on the problem) shown in the diagram in Section 5, where the types in $Z$ are allocated nothing (and pay nothing), the types in $W$ are allocated the grand bundle (at price 12), while the types in $Y$ are allocated item 2 with probability 1 and item 1 with probability 50% (at price 8).
Characterization of Mechanism Optimality. 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 in Theorem 3) obtains necessary and sufficient conditions characterizing the optimality of arbitrary mechanisms with a finite menu size:

Suppose that we are given a feasible mechanism $M$ whose set of possible allocations is finite. We can then partition the type set into finitely many subsets (called regions) $R_1, \ldots, R_k$ of types who enjoy the same price and allocation. The question is this: for what type distributions is $M$ optimal? Theorem 3 answers this question with a sharp characterization result: $M$ is optimal if and only if the measure $\mu$ (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 $\mu$ restricted to region $R_i$ ought to satisfy depends on the allocation to types from $R_i$, 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 domination 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, turning the stochastic domination conditions of Theorem 3 into a standard dual witness for the optimality of a given mechanism, as required by Theorem 2, and conversely showing that a standard witness always implies that the stochastic domination conditions of Theorem 3 hold. We expect that the technical tools used to prove Theorem 3 can be applied to obtain analogous stochastic dominance conditions for even broader classes of mechanisms.

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 measure $\mu$ satisfies a pair of stochastic domination conditions. In particular, if $Z$ are the types who cannot afford the grand bundle and $W$ the types who can, then our equivalent condition to grand-bundling optimality states that $\mu_-$ (the negative part of $\mu$) restricted to $Z$ should convexly dominate $\mu_+$ (the positive part of $\mu$) restricted to $Z$, and $\mu_+$ restricted to $W$ should “second-order stochastically dominate” $\mu_-$ restricted to $W$. Already our characterization of grand-bundling optimality settles a long line of research in the literature, 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 3 and construct intricate couplings establishing the stochastic domination conditions required by our 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 sufficiently large $n$, the optimal mechanism is not (!) a grand-bundling mechanism. See Section 6 for the proofs of these results.

Related Work. There is a rich literature on multi-item mechanism design that pertains 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. Manelli and Vincent [MV06, MV07] provide sufficient conditions for the optimality of grand bundling as well as more complex deterministic mechanisms. An interesting feature of their result is that arbitrarily complex mechanisms may be optimal for certain distributions. Similarly, [DDT13, GK14] provide sufficient conditions for the optimality of general (possibly randomized) mechanisms. All these works use duality theory to establish their results, albeit they relax some of the truthfulness constraints and are therefore only applicable when the relaxed constraints are not binding at the optimum. In particular, all these works on sufficient conditions apply to limited settings of $n$ and $f$. Finally, Rochet and Choné [RC98] study a closely related setting that encompasses the multiple good monopoly when 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 such production function is not strictly convex, 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 [Vil09], 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.

2 Revenue Maximization as an Optimization Program

2.1 Setting up the Optimization Program

Our goal is to find the revenue-optimal mechanism $M$ for selling $n$ goods to a single additive buyer. An additive buyer has a type $x$ specifying his value for each good. The type $x$ 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, we consider direct mechanisms. A (direct) mechanism consists of two functions: (i) an allocation function $P : X \rightarrow [0, 1]^n$ specifying the probabilities, for each possible type declaration of the buyer, that the buyer will be allocated each good, and (ii) a price function $T : X \rightarrow \mathbb{R}$ 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 $x \cdot P(x') - T(x')$.

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

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

When a buyer truthfully reports his type to a mechanism $M = (P, T)$ (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 $M$. It follows by the definitions of $P$ and $T$ that $u(x) = x \cdot P(x) - T(x)$. 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 $\ell_1$ norm and that any utility function satisfying these properties is the utility function of an IC and IR mechanism with $P(x) = \nabla u(x)$ and $T(x) = P(x) \cdot x - u(x)$.

We clarify that a function $u$ is 1-Lipschitz with respect to the $\ell_1$ norm if $u(x) - u(y) \leq \|x - y\|_1$ for all $x, y \in X$. This is essentially equivalent to all partial derivatives having magnitude at most 1 in each dimension.

We will formulate the mechanism design problem 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 $\ell_1$ norm functions $u : X \rightarrow \mathbb{R}$.

In this notation, a mechanism $M$ 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:

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

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

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

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 rewrite the revenue maximization problem as:

\[
\sup_{u \in \mathcal{U}(X) \cap \mathcal{L}_1(X)} \left\{ \int_{\partial X} u(x) f(x) (x \cdot \hat{n}) dx - \int_X u(x) (\nabla f(x) \cdot x + (n+1) f(x)) dx + u(x^{\text{low}}) \right\}
\]

where \( \hat{n} \) denotes the outer unit normal field to the boundary \( \partial X \). To simplify notation we define:

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

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

for all measurable sets \( A \).

With this notation, the above optimization problem becomes \( \sup_{u \in \mathcal{U}(X) \cap \mathcal{L}_1(X)} \int_X u d\mu \). Note that for a constant utility function \( u(x) = 1 \), we have that \( \int_X [\nabla u(x) \cdot x - (u(x) - u(x^{\text{low}}))] f(x) dx = 0 \) and thus we have that \( \mu(X) = \int_X 1 d\mu = 0 \). Furthermore, we have \( |\mu|(X) < \infty \), since \( f, \nabla f \) and \( X \) are bounded.

**Theorem 1.** 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 \( f : X \to \mathbb{R}_{\geq 0} \) is equivalent to solving the optimization problem

\[
\sup_{u \in \mathcal{U}(X) \cap \mathcal{L}_1(X)} \int_X u d\mu
\]

where \( \mu \) is the transformed measure of \( f \) given in (1).

\[\text{It follows from boundedness of } f\text{'s partial derivatives that } \mu \text{ is a Radon measure. Throughout this paper, all "measures" we use will be Radon measures.}\]
2.2 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 \( \mu \) of \( f \) is given by the relation

\[
\mu(A) = I_A(a_1, \ldots, a_n) + \frac{1}{v} \int_{\partial X} I_A(x)(x \cdot \hat{n})dx - \frac{n+1}{v} \int_X I_A(x)dx
\]

for all measurable sets \( A \). Thus, by Theorem 1, the optimal revenue is \( \sup_{u \in \mathcal{U}(X) \cap L_1(X)} \int_X ud\mu \), where \( \mu \) is the sum of:

- A point mass of +1 at the point \((a_1, \ldots, a_n)\).
- 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 \( \{x \in \partial X : x_i = a_i\} \).

3 The Strong Mechanism Design Duality Theorem

Even though the optimal mechanism design problem has a nice formulation as an optimization problem, it is not easy to compute a solution directly. A lot of work has been done on computing an appropriate dual and solving that instead [MV06, DDT13, GK14] but all these approaches are restricted since they only achieve weak duality and work with a relaxed version of the problem. One of our main contributions is to pin down the right dual formulation for the problem and show strong duality. This proves that it is always possible to find a primal/dual pair of solutions to certify optimality and enables us to characterize optimal mechanisms by conditions that are both necessary and sufficient.

Proving such a 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 for the appendix.

3.1 Measure-Theoretic Preliminaries

We first define the following measure-theoretic notation.

**Definition 4.** We define the notation:

Let \( \Gamma(X) \) and \( \Gamma_+(X) \) denote the sets of signed and unsigned (Radon) measures on \( X \).

Consider an unsigned measure \( \gamma \in \Gamma_+(X \times X) \). We denote by \( \gamma_1, \gamma_2 \) the two marginals of \( \gamma \), 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 $\mu$ and a measurable $A \subseteq X$, we define the restriction of $\mu$ to $A$, denoted $\mu|_A$, by the property $\mu|_A(S) = \mu(A \cap S)$ for all measurable $S$.

For a signed measure $\mu$, we will denote by $\mu_+, \mu_-$ the positive and negative parts of $\mu$, respectively. That is, $\mu = \mu_+ - \mu_-$, where $\mu_+$ and $\mu_-$ provide mass to disjoint subsets of $X$.

Our dual problem optimizes over measures satisfying a certain stochastic dominance property. In particular,

**Definition 5.** We say that $\alpha$ convexly dominates $\beta$ for $\alpha, \beta \in \Gamma(X)$, denoted $\alpha \succeq_{\text{cvx}} \beta$, 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 $A \succeq_{\text{cvx}} B$ if $E[u(A)] \geq E[u(B)]$ for all $u \in \mathcal{U}(X)$.

Convex dominance is a weaker condition than standard first-order stochastic dominance, as first-order dominance requires that the inequality holds even for non-convex $u$. For intuition, a measure $\alpha \succeq_{\text{cvx}} \beta$ if we can transform $\beta$ to $\alpha$ 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 6 in Appendix B.1.

### 3.2 Mechanism Design Duality

The main result of this paper is that the mechanism design problem established in Theorem 1 has a dual problem, as follows:

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

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

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

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 $\mu$ into a new measure $\mu'$ with $\mu'(X) = 0$ such that $\mu' \succeq \text{cvx} \mu$. 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 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 $\ell_1$ distance $\|x - y\|_1$, and we are asked for the cheapest method of transforming the positive part of $\mu'$ into the negative part of $\mu'$. Transportation problems of this form have been studied in the mathematical literature. See [Vil09].

So, overall, the goal is to match the positive mass to the negative at a minimum cost where some operations come for free and some come at a cost equal to the distance that the mass was moved.

We remark that one direction of the duality theorem is easy to prove. Proving the reverse direction in Appendix A is significantly more challenging, and relies on non-trivial analytical results such as the Fenchel-Rockafellar duality theorem.

**Lemma 1** (Weak Duality). Let $\mu \in \Gamma(X)$ be the transformed measure of the probability density $f$. Then

$$\sup_{u \in U(X) \cap L^1(X)} \int_X u d\mu \leq \inf_{\gamma \in \Gamma_+(X \times X)} \int_{X \times X} \|x - y\|_1 d\gamma.$$

**Proof of Lemma 1**: For any feasible $u$ for the left-hand side and feasible $\gamma$ 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 $\gamma_1 - \gamma_2 \succeq \text{cvx} \mu$ and the second inequality follows from the 1-Lipschitz condition on $u$.

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

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

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

**Proof of Corollary 1**: The inequalities in the proof of Lemma 1 are tight precisely when both conditions hold.

**Remark 1**. It is useful to geometrically interpret Corollary 1:
1. We view $\gamma^*_1 - \gamma^*_2$ (denote this by $\mu'$) as a “shuffled” $\mu$. Stemming from the $\mu' \succeq_{\text{cvx}} \mu$ constraint, the shuffling of $\mu$ into $\mu'$ is obtained via any sequence of the following operations: (1) Picking a positive point mass $\delta_x$ from $\mu_+$ 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 $u^*(x) = u^*(y)$ for all $z \in \prod_j [x_j, y_j]$. Thus, if $z$ is strictly larger than $x$ in coordinate $i$, then $(\nabla u^*)_i = 0$ at all points “in between” $x$ and $y$. The other operation we are allowed, called a “mean-preserving spread,” is (2) picking a positive point mass $\delta_x$ from $\mu_+$, 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.

2. The second condition is more straightforward than the first. We view $\gamma^*$ as a “transport” map between its component measures $\gamma^*_1$ and $\gamma^*_2$. The condition states that if $\gamma^*$ 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 $\gamma^*$, then they are optimal for their respective problems. This is useful since constructing a $\gamma$ 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 $\gamma^*$ satisfying the conditions of Corollary 1.

Remark 2. Our duality framework, by achieving strong duality, encompasses all prior duality-based frameworks for optimal mechanism design in our setting [MV06, DDT13, GK14]. In particular, if we tighten the $\gamma_1 - \gamma_2 \succeq_{\text{cvx}} \mu$ constraint in the dual problem to a first-order stochastic dominance constraint (maintaining the weak duality property but creating a possible gap between optimal primal and dual values), we essentially recover the duality framework of [DDT13, GK14], which used optimal transport to dualize a relaxed version of the mechanism design problem in which the convexity constraint on $u$ was dropped.

4 Single-Item Applications of Duality, 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 \([\bar{z}, \bar{z}]\). Since \( n = 1 \), if we were to apply our duality framework to this setting, we would choose \( \mu \) according to (1) as follows:

\[
\mu(A) = \mathbb{I}_A(\bar{z}) \cdot (1 - f(\bar{z}) \cdot \bar{z}) + \mathbb{I}_A(\bar{z}) \cdot f(\bar{z}) \cdot \bar{z} - \int_{\bar{z}}^\infty \mathbb{I}_A(z)(f'(z) \cdot z + 2f(z))dz \\
= \mathbb{I}_A(\bar{z}) \cdot (1 - f(\bar{z}) \cdot \bar{z}) + \mathbb{I}_A(\bar{z}) \cdot f(\bar{z}) \cdot \bar{z} - \int_{\bar{z}}^\infty \mathbb{I}_A(z) \left( z - \frac{1 - F(z)}{f(z)} \right) f(z) \right)'dz 
\]

We can interpret the transportation problem of Theorem 2, defined in terms of \( \mu \), as follows:

- 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, \( z \), in the support have an excess supply of \( 1 - f(\bar{z}) \cdot \bar{z} \);
- Finally, the sub-population of buyers at each other type, \( z \), have a demand of

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

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 \( \left( z - \frac{1 - F(z)}{f(z)} \right) f(z) \right)'dz \) 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. While the afore-described solution is just feasible for our transportation problem and may, in principle, not be optimal, it actually is optimal. To see this consider the mechanism that allocates the item to all buyers with non-negative virtual type at a fixed price \( p^* \). It is not hard to see that the resulting utility function \( u \) is feasible when \( F \) is regular: indeed, given that the virtual value function is monotone, the resulting utility function is of the form \( \max \{ z - p^*, 0 \} \), which belongs to \( \mathcal{U}([z, \bar{z}]) \cap \mathcal{L}_1([z, \bar{z}]) \). Moreover, according to Remark 1, this utility function is complementary to the afore-described transportation of mass: 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 \( z < p^* \), \( u' = 0 \) 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 from 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 \( \mu \) can be interpreted as the “negative marginal normalized virtual value,” as it assigns measure \( - \left( z - \frac{1 - F(z)}{f(z)} \right) f(z) \right)'dz \) to the interval \([z, z + dz]\), when \( z \neq \bar{z}, \bar{z} \).

---

\(^3\)We remind the reader that a differentiable distribution \( F \) is regular iff 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.
When $F$ is not regular, the afore-described transportation of mass is not optimal due to the non-monotonicity of the virtual values. In this case, we need to pre-process our measure $\mu$ via mean-preserving spreads, prior to 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.

5 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.

5.1 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) gives the following options:

- Buy any single item for a price of $\frac{2}{3}$
- 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 1 and separated by solid lines.

As a first step, we need to compute the transformed measure $\gamma$ of the uniform distribution on [0, 1]. We have already computed $\gamma$ in Section 2.2. It has a point mass of +1 at $(0, 0)$, a mass of $-3$ distributed uniformly over $[0, 1]$, a mass of +1 distributed uniformly on the top boundary of $[0, 1]$, and a mass of +1 distributed uniformly on the right boundary. Notice that the total net mass is equal to 0 within each region $Z, A, B, W$.

To prove optimality of the MV-mechanism, we will construct an optimal $\gamma^*$ for the dual program of Theorem 2 to match the positive mass $\mu^+$ to the negative $\mu^-$. Our $\gamma^*$ will be decomposed into $\gamma^* = \gamma^Z + \gamma^A + \gamma^B + \gamma^W$ and to ensure that $\gamma^1_1 - \gamma^1_2 \succeq_{cvx} \mu$, we will show that

$$\gamma^1_1 - \gamma^1_2 \succeq_{cvx} \mu | Z; \quad \gamma^1_A - \gamma^1_B \succeq_{cvx} \mu | A; \quad \gamma^1_B - \gamma^1_W \succeq_{cvx} \mu | B; \quad \gamma^1_W - \gamma^1_W \succeq_{cvx} \mu | W.$$ 

We will also show that the conditions of Corollary 1 hold for each of the measures $\gamma^Z, \gamma^A, \gamma^B, \gamma^W$ separately, namely $\int u^* d(\gamma^S_1 - \gamma^S_2) = \int_S u^* d\mu$ and $u^*(x) - u^*(y) = \|x - y\|_1$ hold $\gamma^S$-almost surely for $S = Z, A, B, W$.

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

**Construction of $\gamma^A$ and $\gamma^B$:** In region $A$, $\mu_+|_A$ is distributed on the right boundary while $\mu_-|_A$ is distributed uniformly on the interior of $A$. We construct $\gamma^A$ 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 $\ell_1$ 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 $\gamma^A$ 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 $\gamma^B$ is similar.

**Construction of $\gamma^W$** 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 $u(x) - u(y) = (x_1 + x_2 - \frac{4\sqrt{2}}{3}) - (y_1 + y_2 - \frac{4\sqrt{2}}{3}) = x_1 + x_2 - y_1 - y_2 = \|x-y\|_1$. 

---

Figure 1: The MV-mechanism for two i.i.d. uniform $[0, 1]$ items.
5.2 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 since the distributions are not identical, and thus the characterization of [Pav11] does not apply. In addition, the relaxation-based duality framework of [DDT13] (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 obtaining 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 receives both goods for a price of 12.

This example was constructed for ease of illustration. While our proof only verifies the optimality of our proposed mechanism, in Section 7 we discuss techniques to help us find candidate optimal mechanisms. Drawing inspiration from the work of [Pav11], we expect that optimal mechanisms for two uniform items assign zero utility to a subset $Z$ of types that has pentagonal shape. Our example here is a degenerate one in which only the top edge of the pentagon is non-trivial, resulting in the triangular shape of $Z$.

The proof of optimality works by constructing a measure $\gamma = \gamma^Z + \gamma^Y + \gamma^W$ separately in each region. The constructions of $\gamma^W$ and $\gamma^Z$ are similar to the previous example. The construction of $\gamma^Y$, 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. The proof is presented in the Appendix A.5.

6 Characterization of Optimal Finite-Menu Mechanisms

In the previous section, in order to prove the optimality of a mechanism, we worked by explicitly constructing a measure $\gamma$ separately for each of the regions where items were allocated with different probabilities, so that the conditions of Corollary 1 hold.
In this section, we formalize the previous approach establishing that decomposing the problem into these regions and working on them separately to prove optimality of a mechanism is always possible. We are thus able to obtain conditions which are both necessary and sufficient under which a given mechanism $\mathcal{M} = (\mathcal{P}, \mathcal{T})$ is revenue-optimal. In particular, we show in Theorem 3 that a mechanism is optimal if and only if appropriate stochastic dominance relations are satisfied by the restriction of the transformed measure $\mu$ to each “region” of types, where a region is the collection of types to which the mechanism assigns the same outcome.

To prove this result, we consider the “menu” of different choices given to the buyer by the mechanism:

**Definition 6.** The menu of a mechanism $\mathcal{M} = (\mathcal{P}, \mathcal{T})$ is the set

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

An IC mechanism allocates to every type $x$ a choice in the menu that maximizes that type’s utility. Figure 2 shows an example of a menu and the corresponding set of types that prefer each choice.

![Figure 2: Partition of the type set $X = [0, 100]^2$ induced by some finite menu of lotteries.](image)

The revenue of a mechanism with finite menu size comes from choices in the menu that are bought with strictly positive probability. Even though, the mechanism might contain options in the menu that are only bought with probability 0, we can get another mechanism that gives identical revenue by removing all those options. We call this the **essential form** of a mechanism.
**Definition 7.** 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, but 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 the importance of our result in that we do not merely derive sufficient conditions to certify optimality of mechanisms, as in [MV06] and [DDT13]. Rather we prove that verifying optimality is equivalent to checking some measure-theoretic inequalities, and we cover arbitrary mechanisms with a finite menu-size. The proof of our result is intricate and requires several technical lemmas, which are presented in the appendix. The most important such lemma for our purposes is Lemma 10, which shows that the optimal mass shuffling transforming $\mu$ to $\mu' = \gamma_1 - \gamma_2 \succeq_{\text{cvx}} \mu$, such that the conditions of our strong duality theorem hold, never shuffles mass across regions that allocate the goods with different probabilities. Similarly, we argue that optimal $\gamma$ never transports mass across regions.

Before formally describing our result, it is helpful to provide some intuition behind it. For a region $R$ corresponding to a menu choice $(\vec{p}, t)$ of an optimal mechanism, and given that the corresponding optimal $\gamma$ doesn’t transport mass between regions and that the associated “shuffling” transforming $\mu$ to $\mu'$ likewise doesn’t shuffle across regions, Corollary 1 implies that $\mu_+|_R$ can be transformed to $\mu_-|_R$ using the following 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 $p_i = 0$
- 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 $p_i = 1$

To formally state the theorem, we present it in the form of stochastic conditions which in general are easier to verify than explicitly constructing a transport map $\gamma$. We will need the following definition which extends the notion of convex domination.

**Definition 8.** We say that a function $u : X \to \mathbb{R}$ 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 $v_i = -1$.

We say that $\alpha$ convexly dominates $\beta$ 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 bounded functions $u : \mathbb{R}^n \to \mathbb{R}$:

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

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 bounded functions $u : \mathbb{R}^n \to \mathbb{R}$.
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 the more standard notion of second-order stochastic dominance:

\[
\alpha \succeq_{\text{cvx}(-\vec{1})} \beta \iff \beta \succeq_2 \alpha
\]

We note that two measures satisfying such dominance conditions 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.

**Definition 9 (Optimal Menu Conditions).** We say that a mechanism \( M \) satisfies the optimality conditions with respect to \( \mu \) if for all menu choices \( (p, t) \in \text{Menu}_M \) we have

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

where \( R = \{x \in X : (\mathcal{P}(x), T(x)) = (p, t)\} \) is the subset of types that receive \( (p, t) \) and

\[
v_i = \begin{cases} 
1 & \text{if } p_i = 0 \\
-1 & \text{if } p_i = 1 \\
0 & \text{otherwise}
\end{cases}
\]

**Theorem 3 (Optimal Menu Theorem).** Let \( \mu \) be the transformed measure of a probability density \( f \). Then a mechanism \( M \) 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 \( \mu \).

The proof of the theorem is given in Appendix B.

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 measure \( \mu \) satisfies a pair of stochastic domination 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 some price \( p \) is optimal if and only if the transformed measure \( \mu \) of \( f \) satisfies \( \mu|_W \succeq_2 0 \succeq_{\text{cvx}} \mu|_Z \), where \( W \) is the subset of types that can afford the grand bundle at price \( p \), and \( Z \) the subset of types who cannot.

Next, we explore implications of the characterization of grand bundling optimality.
Example of Grand Bundling Optimality

We now present an example of applying our characterization to prove optimality of a mechanism that makes a take-it-or-leave-it offer for the grand bundle. This result applies to a setting with arbitrarily many items, which is relatively rare in the optimal mechanism design literature.

We consider a setting with \( n \) iid goods whose values are uniform on \((c, c + 1)\). By a simple concentration bound, we can 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. 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.

**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. The proof of this lemma and of Theorem 5 appears in Appendix C.

**Lemma 2.** 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 \( \varphi : A \to B \) 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 \( x_n \geq 1 - \left( \frac{\epsilon n^{-1} + \rho - 1}{\rho} \right)^{1/(n-1)} \).

The main difficulty in proving Theorem 5 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 2 to show a desired stochastic dominance relation holds for an appropriate pairing of regions. The proof of Theorem 5 is in Appendix C.

We now 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 2.2, under the transformed measure \( \mu \) 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 \( 1_{x_1 = c+1}(x) \) (which is increasing and convex in \([c, c+1]^n\)) we have that \( \int_{Z_p} 1_{x_1 = c+1} \, d\mu = \mu(Z_p \cap \{x_1 = c+1\}) = \mu_+(Z_p \cap \{x_1 = c+1\}) > 0 \) 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 $Z_{nc+1}$. The first is equal to:

$$(n + 1) \times \text{Vol} \left\{ x \in (c, c + 1)^n : \|x\|_1 \leq nc + 1 \right\} = (n + 1) \times \frac{n!}{n!}$$

while the latter is equal to:

$$n \times c \times \text{Vol} \left\{ x \in (c, c + 1)^{n-1} : \|x\|_1 + c \leq nc + 1 \right\} = n \times c \times \text{Vol} \left\{ x \in (0, 1)^{n-1} : \|x\|_1 \leq 1 \right\} = \frac{nc}{(n-1)!}$$

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.

\[\Box\]

7 Constructing Optimal Mechanisms

7.1 Preliminaries

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 $\mu_+|_R \preceq_{\text{cvx}(\vec{v})} \mu_-|_R$ 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 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 $\ell_1$ distance between the buyer’s type and the closest point in the exclusion set. (Buyer types within the exclusion set receive zero utility, and hence the name.) All known instances of optimal mechanisms for independently distributed items fall under this category. We proceed to define these concepts formally.

Definition 10 (Exclusion Set). Let $X = \prod_{i=1}^{n} [x_i^{\text{low}}, x_i^{\text{high}}]$. An exclusion set $Z$ of $X$ is a convex, compact, and decreasing subset of $X$ with nonempty interior.
Definition 11 (Mechanism of an Exclusion Set). Every exclusion set \( Z \) of \( X \) induces a mechanism whose utility function \( u_Z : X \to \mathbb{R} \) 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. The proof of the following claim is straightforward casework and appears in Appendix D.

Claim 1. 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 \( \ell_1 \) norm) at most 1. In particular, \( u_Z \) is the utility function of an incentive compatible and individually rational mechanism.

7.2 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 4 is straightforward.

Definition 12 (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 \((x_{\text{crit}}, y_{\text{crit}})\), such that

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

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

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

with domain \([0, x_{\text{crit}}]\) and \([0, y_{\text{crit}}]\) respectively.

Definition 13 (Canonical partition). Let \( Z \) be an exclusion set of \( X \) with critical point \((x_{\text{crit}}, y_{\text{crit}})\) as in Definition 12. We define the canonical partition of \( X \) induced by \( Z \) to be the partition of \( X \) into \( Z \cup A \cup B \cup W \), where

\[
A = \{(x, y) \in X : x < x_{\text{crit}}\} \setminus Z; \quad B = \{(x, y) \in X : y < y_{\text{crit}}\} \setminus Z; \quad W = X \setminus (Z \cup A \cup B),
\]
as shown in Figure 4.

equal to \( b \), if \( b \in Z \) then \( a \in Z \) as well.
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.

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

**Claim 2.** Let $Z$ be an exclusion set of $X$ with outer boundary functions $s_1, s_2$ and critical price $P$, and let $Z \cup A \cup B \cup W$ 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:

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

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

Pick any $v = (v_1, v_2) \in A$. 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, $\max_{x \in [0, v_1]}(x + s_1(x))$ is maximized at $x = v_1$, we have that, $\|v - z'||_1 \geq \|v||_1 - v_1 - s_1(v_1) = v_2 - s_1(v_1) = \|v - z^*||_1$.

We now describe sufficient conditions under which $u_Z$ is optimal.

**Definition 14 (Well-formed canonical partition).** Let $Z \cup A \cup B \cup W$ be a canonical partition of $X$ induced by exclusion set $Z$ and let $\mu$ be a signed Radon measure on $X$ such that $\mu(X) = 0$. We say that the canonical partition is well-formed with respect to $\mu$ if the following conditions are satisfied:
1. \( \mu|_Z \preceq \text{ex} \) 0 and \( \mu|_W \succeq 2 \), and

2. for all \( v \in X \) and all \( \epsilon > 0 \):
   - \( \mu|_A([v_1, v_1 + \epsilon] \times [v_2, \infty)) \geq 0 \), with equality whenever \( v_2 = 0 \)
   - \( \mu|_B([v_1, \infty) \times [v_2, v_2 + \epsilon]) \geq 0 \), 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 \( W \). We interpret Condition 2 as saying that \( \mu|_A \) (resp. \( \mu|_B \)) 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 \( \mu \) 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 4 to visualize such a mechanism.

**Theorem 7.** Let \( \mu \) be the transformed measure of a probability density function \( f \). If there exists an exclusion set \( Z \) inducing a canonical partition \( Z \cup A \cup B \cup W \) of \( X \) that is well-formed with respect to \( \mu \), 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 \( (x, y) \in A \), 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 \( (x, y) \in B \), 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 \( (x, y) \in W \), 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 12.

**Proof.** We will show that \( u_Z \) maximizes \( \sup_{u \in \mathcal{U} \cap C_1(X)} \int_X u \, d\mu \). By Corollary 1, it suffices to provide a \( \gamma \in \Gamma_+(X \times X) \) such that \( \gamma_1 - \gamma_2 \preceq \text{ex} \) \( \mu \), \( \int u_X d(\gamma_1 - \gamma_2) = \int u_Z d\mu \), and \( u_Z(x) - u_Z(y) = \|x - y\|_1 \) holds \( \gamma \)-almost surely. The \( \gamma \) we construct will never transport mass between regions. That is, \( \gamma = \gamma_Z + \gamma_W + \gamma_A + \gamma_B \) where\(^6\)

\(^6\)We chose this notation for simplicity, where \( \gamma_Z \in \Gamma_+(Z \times Z) \), \( \gamma_W \in \Gamma_+(W \times W) \), and so on.
• $\gamma_Z = 0$. We notice that $(\gamma_Z)_1 - (\gamma_Z)_2 = 0 \succeq_{\text{cvx}} \mu|_Z$.

• $\gamma_W$ is constructed such that $(\gamma_W)_1 - (\gamma_W)_2 \succeq_{\text{cvx}} \mu|_W$ and the component-wise inequality $x \geq y$ holds $\gamma_W(x,y)$ almost surely.\(^7\) As in our proof of Theorem 3, the existence of such a $\gamma_W$ is guaranteed by Strassen’s theorem for second order dominance (see Lemma 6 in the appendix).

• $\gamma_A \in \Gamma_+(A \times A)$ will be constructed to have respective marginals $\mu_+|_A$ and $\mu_-|_A$, and so that, $\gamma_A(x,y)$ almost surely, it holds that $x_1 = y_1$ and $x_2 \geq y_2$. Thus, $(\gamma_A)_1 - (\gamma_A)_2 = \mu|_A$, and $\gamma_A$ sends positive mass “downwards.”\(^8\) We claim that such a map can indeed be constructed, by noticing that Property 2 of Definition 14 guarantees that, restricted to any vertical strip inside $A$, $\mu_+$ first-order stochastically dominates $\mu_-$.\(^9\) Hence, Strassen’s theorem for first-order dominance guarantees that restricted to that strip $\mu_+$ can be coupled with $\mu_-$ so that, with probability 1, mass is only moved downwards.

Measure $\gamma_A$ satisfies $x_1 = y_1$, $\gamma_A(x,y)$ almost surely, and hence also

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

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

It follows by our construction that $\gamma = \gamma_Z + \gamma_W + \gamma_A + \gamma_B$ satisfies all necessary properties to certify optimality of $u_Z$.

\[\square\]

8 Applying Theorem 7 to Construct Optimal Mechanisms

In this section, we provide example applications of Theorem 7. A technical difficulty is verifying the stochastic dominance relation $\mu|_W \succeq_2 0$ required to apply the theorem. In our examples, we will have the stronger condition $\mu|_W \succeq_1 0$, which is easier to verify, yet still imposes technical difficulties. In Section 8.1, we present a useful tool, Lemma 3, for verifying first-order stochastic dominance. In Section 8.2 we then provide example applications of Theorem 7 and Lemma 3 to solve for optimal mechanisms.

8.1 Verifying First-Order Stochastic Dominance

The following lemma presents a useful tool for verifying first order stochastic dominance between measures.\(^{10}\)

---

\(^7\)As in Example 2 and as discussed in Remark 1, we aim for $\gamma_W$ to transport “downwards and leftwards” since both items are allocated with probability 1 in $W$.

\(^8\)Once again, the intuition for this construction follows Remark 1.

\(^9\)Indeed, as $\epsilon \to 0$, Property 2 states exactly the one-dimensional equivalent condition for first-order stochastic dominance in terms of cumulative density functions.

\(^{10}\)The lemma also appeared as Theorem 7.4 of [DDT13] without a proof. We provide a detailed proof in Appendix E.1.
Lemma 3. Let $C = [p_1, q_1] \times [p_2, q_2]$ where $q_1$ and $q_2$ are possibly infinite and let $R$ be a decreasing nonempty subset of $C$. Consider two measures $\kappa, \lambda \in \Gamma_+(C)$ with bounded integrable density functions $g, h : C \to \mathbb{R}_{\geq 0}$ respectively that satisfy the following conditions.

- $g(x, y) = h(x, y) = 0$ for all $(x, y) \in R$.
- $\int_C g(x, y) dx dy = \int_C h(x, y) dx dy < \infty$.
- 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 $\alpha : [p_1, q_1] \to \mathbb{R}_{\geq 0}$ and $\beta : [p_2, q_2] \to \mathbb{R}_{\geq 0}$, and an increasing function $\eta : C \to \mathbb{R}$ such that for all $(x, y) \in C \setminus R$:
  \[ g(x, y) - h(x, y) = \alpha(x) \cdot \beta(y) \cdot \eta(x, y) \]

Then $\kappa \succeq_{1} \lambda$.

This result provides a sufficient condition for a measure to stochastically dominate another in the first order. A complete proof of Lemma 3 is in Appendix E.1 and is an application of Claim 13, also found in Appendix E.1, 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 3 are satisfied, we can induct on the number of boxes by removing one box at a time. We note that Lemma 3 is applicable even to distributions with unbounded support.

Lemma 3 deals with the scenario where two density functions, $g$ and $h$, are both nonzero on some set $C \setminus R$, where $R$ is a decreasing subset of $C$. This setup is motivated by Figure 4. Recall that, in order to apply Theorem 7, we need to check a second order stochastic dominance condition in region $W$, namely $\mu|_W \succeq_{2} 0$. Instead, it suffices to show the first order stochastic dominance $\mu|_W \succeq_{1} 0$, which we plan to show via Lemma 3 by taking $C = [x_{\text{crit}}, \infty) \times [y_{\text{crit}}, \infty)$, $R = C \cap Z$, and $g, h$ the densities corresponding to measures $\mu_+|_W$ and $\mu_-|_W$. To prove that $g$ dominates $h$ in the first order, Lemma 3 states that it suffices to verify that (1) $g - h$ has an appropriate form; (2) the integral of $g - h$ on $C$ is zero; and (3) if we integrate $g - h$ along either a vertical or horizontal line outwards starting from any point in $R$, the result is negative.

8.2 Examples

We apply Theorem 7 to obtain optimal mechanisms in several two-item settings. In Section 8.2.1, we consider two independent items distributed according to beta distributions. We find the optimal mechanism, showing that it actually offers an uncountably infinite menu of lotteries. We conclude with Section 8.2.2 where we discuss extensions of Theorem 7 to distributions with infinite support,
providing the optimal mechanism for two arbitrary independent exponential items, as well as the optimal mechanism for an instance with two independent power-law items.

### 8.2.1 An Optimal Mechanism with Infinite Menu Size: Two Beta Items

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 $S_{\text{top}}$ and $S_{\text{right}}$, (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 8.2.1—8.2.2, 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.

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 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 $\mu$ induced by $f$. Notice that

$$-\nabla f(x,y) \cdot (x,y) - 3f(x,y)$$

$$= -xf_2(y)\frac{\partial f_1(x)}{\partial x} - yf_1(x)\frac{\partial f_2(y)}{\partial y} - 3f_1(x)f_2(y)$$

$$= -xf_2(y)((a_1-1)x^{a_1-2}(1-x)^{b_1-1} - (b_1-1)x^{a_1-1}(1-x)^{b_1-2})$$

$$+ yf_1(x)((a_2-1)y^{a_2-2}(1-y)^{b_2-1} - (b_2-1)y^{a_2-1}(1-y)^{b_2-2})$$

$$- 3f_1(x)f_2(y)$$

$$= -(a_1-1)f_1(x)f_2(y) + (b_1-1)\frac{x}{1-x}f_1(x)f_2(y)$$

$$- (a_2-1)f_1(x)f_2(y) + (b_2-1)\frac{y}{1-y}f_1(x)f_2(y) - 3f_1(x)f_2(y)$$

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

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 $\mu$ 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 \(x_i = 1\).

Concrete Example  
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 \(\mu\) 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 \(\mu\) is positive on \(P = \{(x,y) \in (0, 1)^2 : \frac{1}{1-x} + \frac{1}{1-y} > 5\} \cup \{\vec{0}\}\) and non-positive on \(N = \{(x,y) \in [0, 1)^2 \setminus \{\vec{0}\} : \frac{1}{1-x} + \frac{1}{1-y} \leq 5\}\), and that \(N \cup \{\vec{0}\}\) 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 \(\mu\) “upwards” from \(t = y\) to \(t = 1\) yields zero. Since \(N \cup \{\vec{0}\}\) is a decreasing set, it follows that \(S_{\text{top}} \subset N\) and that integrating \(\mu\) upwards starting from any point above \(S_{\text{top}}\) 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 \(S_{\text{right}} \subset N\). \(S_{\text{top}}\) and \(S_{\text{right}}\) are shown in Figure 5.

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 \(S_{\text{top}}\) and the region below \(S_{\text{right}}\) 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 \(p^* \approx 0.5535\) as the intercept of the 45° line in Figure 5 which causes \(\mu(Z) = 0\)
for the set $Z \subset [0,1]^2$ lying below $S_{\text{top}}$, $S_{\text{right}}$ and the 45° line. We can also compute the critical point $(x_{\text{crit}}, y_{\text{crit}}) \approx (0.0618, 0.0618)$ by finding the intersection of the critical price line with the sets $S_{\text{top}}$ and $S_{\text{bottom}}$. Moreover, by the definition of the sets $S_{\text{top}}$ and $S_{\text{bottom}}$, 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:

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

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

as illustrated in Figure 5.

![Figure 5: The well-formed canonical partition for $f_1(x) = 2(1-x)$ and $f_2(y) = 2(1-y)$.](image)

**Step (iv).** We claim that the canonical partition $Z \cup A \cup B \cup W$ is well-formed with respect to $\mu$. Condition 2 is satisfied by construction of $S_{\text{top}}$ and $S_{\text{right}}$ 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 $R = Z, A, B$ and $W$, 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, $\mu$ has negative density along these curves and all points below either curve, other than at the origin. Hence, $\mu \mid_Z \geq_1 \mu \mid_{Z_{\text{crit}}} \text{ which implies that } \mu \mid_Z \geq_0 0$. Hence, the only non-trivial condition of Definition 14 that we need to verify is $\mu \mid_W \geq_2 0$. In fact, we can apply Lemma 3 to conclude the stronger dominance relation $\mu \mid_W \geq_1 0$. See Appendix E.2. Having verified all conditions of Definition 14 we can now apply Theorem 7 to conclude 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 \((x, y) \in A\), the buyer receives item 1 with probability \(-s_1'(x) = \frac{2}{(4-5x)^2}\), item 2 with probability 1, and is charged \(\Delta = s_1(x) - xs_1'(x) = \frac{2-3x}{4-5x} + \frac{2x}{(4-5x)^2}\).

- If \((x, y) \in B\), the buyer receives item 2 with probability \(-s_2'(y) = \frac{2}{(4-5y)^2}\), item 1 with probability 1, and is charged \(\Delta = s_2(y) - ys_2'(y) = \frac{2-3y}{4-5y} + \frac{2y}{(4-5y)^2}\).

- If \((x, y) \in W\), the buyer receives both items and is charged \(p^* \approx .5535\).

Example 3 establishes that an optimal mechanism might offer 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 many cases where the values for the two items are distributed independently according to different Beta distributions. Figure 6 summarizes the optimal mechanisms for several examples.

Figure 6: Canonical Partitions for different cases of Beta distributions. The shaded region is where the measure \(\mu\) becomes negative. (Note that in the case \(b_i = 1\), \(\mu\) 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), (4) Beta(2,2) and Beta(3,6), (5) Beta(2,3) and Beta(2,2), (6) Beta(6,2) and Beta(9,3)
8.2.2 Distributions of Unbounded Support: Exponential and Power-Law

So far, this paper has focused on type distributions with bounded support. In this section, we note that Theorem 1, Lemma 1, and Theorem 7 can be easily modified to accommodate settings with unbounded type spaces, as long as the type distribution decays sufficiently rapidly towards infinity. On the other hand, we do not know extensions of our strong duality theorem (Theorem 2), and our equivalent conditions for optimal menus (Theorem 3) for unbounded type distributions, due to technical issues.

In Appendix F, we provide a short discussion of the modifications required to obtain an analog of Theorem 7 for unbounded distributions that are sufficiently fast-decaying, and present below two example settings that can be analyzed using the modified characterization theorem. Both examples are taken from [DDT13].

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 $A$ and $B$ are empty), and establishing the optimality of the mechanism amounts to establishing that the measure $\mu$ induced by the type distribution first-order stochastically dominates the $0$ measure in region $W$.

**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 $p^* \approx .35725$.

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 $A$, and possibly region $B$ (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 $\lambda_1$ and $\lambda_2$ offers the following menu:

1. receive nothing, and pay $0$;
2. receive the first item with probability $1$ and the second item with probability $\lambda_2/\lambda_1$, 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

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

where $\mu$ is the transformed measure of the joint distribution.

31
Figure 7: The canonical partition of $\mathbb{R}_{\geq 0}^n$ for the proof of Example 5. In this diagram, $p^* > 2/\lambda_1$. If $p^* \leq 2/\lambda_1$, $B$ is empty. The positive part $\mu_+$ of $\mu$ is supported inside $\mathcal{P} \cap \{\vec{0}\}$ while the negative part $\mu_-$ is supported within $Z_{p^*} \cup \mathcal{N}$.

References


A Strong Mechanism Design Duality - Proof of Theorem 2

In this section, we give a formal proof of the strong mechanism duality theorem. 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 $u$.

A.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 [Vil09], 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 15 (Legendre-Fenchel Transform).** Let $E$ be a normed vector space and let $\Lambda : E \to \mathbb{R} \cup \{+\infty\}$ be a convex function. The Legendre-Fenchel transform of $\Lambda$, denoted $\Lambda^*$, is a map from the topological dual $E^*$ of $E$ to $\mathbb{R} \cup \{\infty\}$ given by

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

**Claim 3 (Fenchel-Rockafellar duality).** Let $E$ be a normed vector space, $E^*$ its topological dual, and $\Theta, \Xi$ two convex functions on $E$ taking values in $\mathbb{R} \cup \{+\infty\}$. Let $\Theta^*, \Xi^*$ be the Legendre-Fenchel transforms of $\Theta$ and $\Xi$ respectively. Assume that there exists $z_0 \in E$ such that $\Theta(z_0) < +\infty$, $\Xi(z_0) < +\infty$ and $\Theta$ is continuous at $z_0$. Then

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

**Lemma 4.** Let $X$ be a compact convex subset of $\mathbb{R}^n$, and let $\mu \in \Gamma(X)$ be such that $\mu(X) = 0$. Then

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

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

**Proof of Lemma 4:** 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 $E^* = \Gamma(X \times X)$.  

---

11 More formally, this follows from Riesz representation theorem.
We now define functions $\Theta, \Xi$ 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 $\Xi$ 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 $\psi'$ differs from $\psi$ only by an additive constant, and $\phi$ differs from $\phi'$ by the same additive constant, and therefore (since $\mu_+$ and $\mu_-$ 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 $\Xi$ is convex, since we can take convex combinations of the $\psi$ and $\phi$ 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 $\Theta$ is continuous at $z_0$. We can thus apply the Fenchel-Fockafellar duality theorem.

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

$$
\Theta^*(-\gamma) = \sup_{f \in CB(X \times X)} \left[ \int_{X \times X} f(x,y) d(-\gamma(x,y)) \right]
$$

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

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

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

The claim follows, since \(k > 0\) is arbitrary.

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

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

\[
= \sup_{\psi, \phi \in \mathcal{U}(X)} \left[ \int_{X \times X} (\psi(y) - \phi(x)) d\gamma(x, y) - \int_X \psi d\mu_- + \int_X \phi d\mu_+ \right]
\]

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 \(\gamma \not\in \Gamma_+(X \times X)\). Moreover, when \(\gamma \in \Gamma_+(X \times X)\):

\[
\Xi^*(\gamma) = \sup_{\psi, \phi \in \mathcal{U}(X)} \left[ \int_{X \times X} (\psi(y) - \phi(x)) d\gamma(x, y) - \int_X \psi d\mu_- + \int_X \phi d\mu_+ \right]
\]

\[
= \sup_{\psi, \phi \in \mathcal{U}(X)} \left[ \int_X \psi d(\gamma_2 - \mu_-) + \int_X \phi d(\mu_+ - \gamma_1) \right]
\]

\[
= \begin{cases} 
0 & \text{if } \gamma_1 \succeq_{cx} \mu_+ \text{ and } \gamma_2 \preceq_{cx} \mu_- \\
\infty & \text{otherwise.}
\end{cases}
\]

The last equality is true because if \(\gamma_1 \succeq_{cx} \mu_+\) 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 \(\phi\) arbitrarily, we can make the inside quantity as large as we want. The same holds when \(\mu_- \not\preceq_{cx} \gamma_2\).

\[\text{Formally, we have used Lusin’s theorem to find such a } g \text{ which is continuous, as opposed to merely measurable.}\]
We now apply Fenchel-Rockafellar duality:

\[
\inf_{f \in CB(X \times X)} \{ \Theta(f) + \Xi(f) \} = \max_{\gamma \in \Gamma(X \times X)} \{ -\Theta^*(-\gamma) - \Xi^*(\gamma) \}
\]

\[
\inf_{f(x,y) \geq \|x-y\|_1} \left( \int_X \psi d\mu_+ - \int_X \phi d\mu_- \right) = \max_{\gamma \in \Gamma_+(X \times X)} \left( -\int_{X \times X} \|x-y\|_1 d\gamma(x,y) - \Xi^*(\gamma) \right)
\]

\[
\sup_{\psi, \phi \in U(X)} \left( \int_X \phi d\mu_+ - \int_X \psi d\mu_- \right) = \min_{\gamma \in \Gamma_+(X \times X)} \left( \int_{X \times X} \|x-y\|_1 d\gamma(x,y) \right).
\]

\[\square\]

A.2 From Two Convex Functions to One

Lemma 5. 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

\[
\sup_{\psi, \phi \in U(X)} \left( \int_X \phi d\mu_+ - \int_X \psi d\mu_- \right) = \sup_{u \in U(X) \cap L_1(X)} \left( \int_X ud\mu_+ - \int_X ud\mu_- \right).
\]

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

Proof of Lemma 5: Given any feasible \( u \) for the right-hand side of Lemma 5, 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 \( \phi \) and \( \psi \) be feasible for the left-hand side. Given \( \phi \), it is clear that \( \psi \) must satisfy \( \psi(y) \geq \sup_x [\phi(x) - \|x-y\|_1] \).

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

Claim 4. \( \tilde{\psi} \in U(X) \) and \( \tilde{\psi} \in L_1(X) \).

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

- **Continuity.** Continuity of \( \tilde{\psi} \) follows from the Maximum Theorem since both \( \phi \) 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 \( \phi \) 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 \( \tilde{\psi} \in U(X) \). We will now show that \( \tilde{\psi} \in L_1(X) \). We have

\[
\tilde{\psi}(x) - \tilde{\psi}(y) = \sup_{z} \inf_{w} (\phi(z) - \|z - x\|_1 - \phi(w) + \|w - y\|_1) \\
\leq \sup_{z} (\phi(z) - \|z - x\|_1 - \phi(z) + \|z - y\|_1) \\
= \sup_{z} (\|z - y\|_1 - \|z - x\|_1) \leq \|x - y\|_1.
\]

\( \square \)

Since \( \phi, \tilde{\psi} \) are a feasible pair of functions for the left-hand side of Lemma 5, we know that \( \phi \) satisfies the inequality \( \phi(x) \leq \inf_y [\tilde{\psi}(y) + \|x - y\|_1] \). We now set \( \bar{\phi}(x) = \inf_y [\tilde{\psi}(y) + \|x - y\|_1] \). It is clear that the value of the left-hand objective function under \( \bar{\phi}, \tilde{\psi} \) is at least as large as its value under \( \phi, \tilde{\psi} \).

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

\[
\bar{\phi}(x) = \inf_y [\tilde{\psi}(y) + \|x - y\|_1] = \tilde{\psi}(x) + \inf_y [\tilde{\psi}(y) - \tilde{\psi}(x) + \|x - y\|_1] \geq \tilde{\psi}(x)
\]
where the last inequality holds since \( \bar{\psi}(x) - \bar{\psi}(y) \leq \|x - y\|_1 \). Therefore \( \bar{\phi} = \bar{\psi} \), and thus \( \bar{\phi} \in U(X) \). 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 5, and the value of the right-hand objective under \( \bar{\phi} \) is at least as large the value of the left-hand objective under \( \phi, \psi \). We notice finally that if \( \phi, \psi \) are optimal for the left-hand side, then \( \bar{\phi} \) is optimal for the right-hand side. \( \square \)

A.3 Proof of Theorem 2

By combining Lemma 1, Lemma 4, and Lemma 5, we have

\[
\inf_{\gamma \in \Gamma_+(X \times X)} \int_{X \times X} \|x - y\|_1 d\gamma \geq \sup_{u \in U(X) \cap L_1(X)} \int_X ud\mu \geq \inf_{\gamma \in \Gamma_+(X \times X)} \int_{X \times X} \|x - y\|_1 d\gamma \left( x, y \right).
\]

By Lemma 4, the last minimization problem above achieves its infimum for some \( \gamma^* \). We notice that \( \gamma^* \) is also feasible for the first minimization problem above, and therefore the inequality is actually an equality and \( \gamma^* \) is optimal for the first minimization problem. In addition, since \( \gamma^* \) is feasible for the last minimization problem, it satisfies \( \gamma_1^*(X) = \gamma_2^*(X) = \mu_+(X) \). All that remains is to prove that the supremum to the maximization problem is achieved for some \( u^* \). A proof of this fact is in Appendix A.4.

A.4 Existence of Optimal Mechanism

We now prove that the supremum of the maximization problem of Theorem 2 is achieved for some \( u^* \). Consider a sequence of feasible functions \( u_1, u_2, \ldots \in U(X) \cap L_1(X) \) such that \( \int_X u_i d\mu \) converges monotonically to the supremum value \( V \), which we have proven is finite.\(^{13}\) 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.

A.5 Omitted Proofs from Section 5 - Example 2

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 \( \mu \) of the type distribution is composed of:

\(^{13}\)Finiteness is also obvious because \( X \) is bounded and the infimum problem is feasible.
• 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 $+\frac{7}{50}$)
• 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 $+\frac{4}{3}$)
• Mass $-\frac{1}{3}$ distributed on left edge of rectangle (Linear density $-\frac{1}{9}$)

We claim that $\mu$ with the remaining negative mass arbitrarily. Noticing that by moving mass downwards, then we match the $\mu$.

We will also show that the conditions of Corollary 1 hold for each of the measures $\gamma$, $\gamma^Y$, and $\gamma^W$. To ensure that$\gamma^A_1 - \gamma^A_2 \succeq_{cvx} \mu$, we will show that $\gamma^Z_1 - \gamma^Z_2 \succeq_{cvx} \mu|Z; \gamma^Y_1 - \gamma^Y_2 \succeq_{cvx} \mu|Y; \gamma^W_1 - \gamma^W_2 \succeq_{cvx} \mu|W$.

We will also show that the conditions of Corollary 1 hold for each of the measures $\gamma^Z$, $\gamma^Y$, and $\gamma^W$ separately, namely $\int u^*d(\gamma^A_1 - \gamma^A_2) = \int_A u^*d\mu$ and $u^*(x) - u^*(y) = \|x - y\|_1$ hold $\gamma^A$-almost surely for $A = Z$, $Y$, and $W$.

Construction of $\gamma^Z$. 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 $\mu|Z \succeq_{cvx} 0$. We therefore set $\gamma^Z$ to be the zero measure, and the relation $\gamma^Z_1 - \gamma^Z_2 = 0 \succeq_{cvx} \mu|Z$, as well as the two necessary equalities from Corollary 1, are trivially satisfied.

Construction of $\gamma^W$. 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.\textsuperscript{14} 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 $\mu_+$ mass on $[8, 16] \times \{7\}$ with the $\mu_-$ mass on $[8, 16] \times \{14\}$ by moving mass downwards, then we match the $\mu_+$ mass on $\{16\} \times [4, \frac{14}{3}]$ with the $\mu_-$ mass on $[\frac{22}{3}, 16] \times (4, \frac{14}{3})$ by moving mass to the left, and finally match the $\mu_+$ mass on $\{16\} \times \{14\}$ 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 $\gamma^Y$. 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

\textsuperscript{14}To 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.
\( \mu_-|_Y \) 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 \( \gamma^Y \) 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 \( \alpha = \alpha_+ - \alpha_- \geq_{\text{cex}} 0 \). As discussed above, we set \( \alpha \) to have density function

\[
f_\alpha(z_1, z_2) = \mathbb{I}_{z_2=7} \cdot \left( \frac{1}{6} \mathbb{I}_{z_1=4} + \frac{1}{24} \left( z_1 - \frac{20}{3} \right) \right) \cdot \mathbb{I}_{z \in Y}.
\]

The measure \( \alpha \) is supported on the line \([4, 8] \times \{7\}\) and consists of a point mass of \( \frac{1}{3} \) 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 \( \alpha \geq_{\text{cex}} 0 \), which we need for feasibility, and that \( \int_Y u^* d\alpha = 0 \), 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 \( \gamma^Y \) 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 \( \gamma^Y \) 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,

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

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 \( 2 \epsilon \cdot \left( \frac{24}{2} - 1 \right) \):

\[
(\mu_-|_Y - \alpha|_Y)([z_1 - \epsilon, z_1 + \epsilon] \times [4, 7])
\]

\[
= \frac{1}{12} \cdot \left( 2 \epsilon \cdot \left( \frac{z_1}{2} - 1 \right) \right) - \frac{1}{24} \int_{z_1-\epsilon}^{z_1+\epsilon} (z - \frac{20}{3}) dz
\]

\[
= \frac{\epsilon z_1}{12} - \frac{\epsilon}{6} - \frac{1}{24} (2\epsilon z_1 - \frac{40\epsilon}{3}) = \frac{7\epsilon}{18} = \mu_+|_Y([z_1 - \epsilon, z_1 + \epsilon] \times [4, 7]).
\]

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

\[15\text{Since } \alpha \text{ is supported on a 1-dimensional line, this verification uses a property analogous to the standard characterization of one-dimensional second-order stochastic dominance via the cumulative density function. Informally, we can argue that } \alpha \geq_{\text{cex}} 0 \text{ by considering integrals of one-dimensional test functions (by restricting our attention to the line } z_2 = 7 \text{) and noticing that, since } \alpha(Y) = 0, \text{ we need only consider test functions } h \text{ which have value 0 at } z_1 = 4. \text{ We then use the fact that all linear functions integrate to 0 under } \alpha \text{ and that (ignoring the point mass at } z_1 = 4, \text{ since } h \text{ is 0 at this point) the density of } \alpha \text{ is monotonically increasing.} \]
second good is received with probability 1), it follows that $\gamma^Y$ satisfies the necessary conditions of Corollary 1.

B Proof of Stochastic Conditions of Section 6

Our goal in this section is to prove Theorem 3. We begin by presenting some useful probabilistic tools that will be essential for the proof.

B.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 6** (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 $E[\hat{B} | \hat{A}] \geq \hat{A}$ 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 8.

**Lemma 7** (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 8** (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 $[0, M]^n \to \mathbb{R}$. Then $E[u(V)] \geq u(E[V]).$ Furthermore, equality holds if and only if, for every $a$ in the subdifferential of $u$ at $E[V]$, the equality $u(V) = a \cdot (V - E[V]) + u(E[V])$ holds almost surely.

The following lemma is a conditional variant of Lemma 8, 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 10, the main result of this subsection.
Lemma 9. Let $(\Omega, \mathcal{A}, P)$ be a probability space, $V$ be a random variable on $\Omega$ 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-$\sigma$-algebra of $\mathcal{A}$ and suppose that $E[u(V)] = u(E[V|\mathcal{C}])$ 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\(^{16}$ $P_{V|\mathcal{C}}(\cdot, x)$, 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 9: 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, $E[V|\mathcal{C}] \in \mathcal{C}$ and $E[u(V)|\mathcal{C}] \geq u(E[V|\mathcal{C}])$. The proof of Theorem 10.2.7 in [Dud02] furthermore shows that the following two equalities hold:

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

Since $E[u(V)|\mathcal{C}](x) = u(E[V|\mathcal{C}](x))$ for almost all $x$, we apply the unconditional Jensen inequality (Lemma 8) to the laws $P_{V|\mathcal{C}}(\cdot, x)$ to prove the lemma. \hfill \Box

We now present Lemma 10. 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 10. 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 $\ell_1$ norm, convex, and monotonically non-decreasing. Suppose that $E[u(A)] = E[u(B)]$ 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 $\hat{B}$.
- $E[\hat{B}|\hat{A}]$ is componentwise greater or equal to $\hat{A}$
- $u(E[\hat{B}|\hat{A}]) = u(\hat{A})$.

Proof of Lemma 10: By Lemma 6, there exist random variables $\hat{A} =_{st} A$ and $\hat{B} =_{st} B$ such that $E[\hat{B}|\hat{A}]$ is componentwise greater than or equal to $\hat{A}$ almost surely. We have

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

\(^{16}\)The law $P_{V|\mathcal{C}}(\cdot, x)$ allows us to express the conditional distribution of $V$ given $\mathcal{C}$.
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 $z \in X$,

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

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

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

Since $\mathbb{E}[u(\hat{B})|\hat{A}] = \mathbb{E}[u(\mathbb{E}[\hat{B}|\hat{A}])]$, 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 9, it therefore holds for almost all values $x$ of $\hat{A}$ that the equality

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

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

$$u(p) \geq u(x) + a_x \cdot (p - x) = u(x) + a_x \cdot (\hat{B}|\hat{A} = x - x) + a_x \cdot (p - \hat{B}|\hat{A} = x)$$

$$= u(\hat{B}|\hat{A} = x) + a_x \cdot (p - \hat{B}|\hat{A} = x).$$

\[ \square \]

### B.2 Proof of the Optimal Menu Theorem (Theorem 3)

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

#### B.2.1 Sufficiency Conditions

We will show that the Optimal Menu Conditions of Definition 9 imply that a mechanism $\mathcal{M}$ is optimal. To show the theorem, we construct a measure $\gamma$ 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 \((p, t) \in \text{Menu}_M\), the corresponding region \(R\) and the corresponding vector \(\vec{v}\) as in Definition 9. 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 \(A|R \preceq_{\text{cvx}} B|R\) (almost surely). By the extended version of Strassen’s theorem (Lemma 7), it holds that there exist random vectors \(\hat{A}, \hat{B}\) with \(\hat{A} =_s A|R\) and \(\hat{B} =_s B|R\), 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 \(\gamma_R\) be the measure according to which the vector \((\hat{A}, \hat{C})\) is distributed. By construction, \(\gamma_{R1} = \mu_+|R\) and \(\gamma_{R2} \preceq_{\text{cvx}} \mu_-|R\), and thus \(\gamma_{R1} - \gamma_{R2} \preceq_{\text{cvx}} \mu|_R\). 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 \(\gamma\) satisfying the conditions of Corollary 1 by considering the sum of the constructed measures for each region.

**B.2.2 Optimality implies Stochastic Conditions**

We will now prove the other direction of the result. Consider an optimal mechanism \(M = (\mathcal{P}, \mathcal{T})\) with a finite menu size over type space \(X = \prod_{i=1}^n[x_i^{\text{low}}, x_i^{\text{high}}]\). Since \(M\) is given in essential form, in the menu of \(M\) 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 11.** \( \mu_-(Z) = 0 \)

**Proof.** Note that, by its construction, \( \mu_- \) 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 \( \mu_- \) assigns no mass to the subset of \( Z \) which lies on the interior of \( X \).

Regarding the \( \mu_- \)-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 \( \mu_- \) 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 \( \gamma \) be the optimal measure of Theorem 2. Then, \( \gamma \) satisfies the properties of Corollary 1. In particular, it holds that:

1. \[
\int ud(\gamma_1 + \mu_-) = \int ud(\mu_+ + \gamma_2)
\]

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 \( \gamma \)) 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

\[
\int (u(x) - \sum_i x_i) d\gamma_1 = \int (u(y) - \sum_i y_i) d\gamma_2
\]

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

We are now ready to use Lemma 10 which follows from Jensen’s inequality. We will apply it in two different steps, which we will then combine to show that \( \mu_+|_R \succeq_{\text{cvx}(\gamma)} \mu_-|_R \).

**Step (ia):** We will first apply Lemma 10 to random variables \( A, B \) distributed according to the measures \( \gamma_2 + \mu_+ \) and \( \gamma_1 + \mu_- \) respectively. Since \( \mu_+ - \mu_- \succeq_{\text{cvx}} \gamma_1 - \gamma_2 \), by the feasibility of \( \gamma \), we have that \( A \succeq_{\text{cvx}} B \). Moreover, \( \mathbb{E}[u(A)] = \mathbb{E}[u(B)] \), from Equation (2) above, and \( u \) is convex and non-decreasing, from the feasibility of \( u \).
To apply Lemma 10, we choose the function $g(x)$, which is a subgradient function 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 $g(x) = P(x)$.
- For all other $x$, $u$ has a continuum of different subgradients at $x$. In particular, any vector in the convex hull of $\{p : p \cdot x - t = u(x), (p, t) \in \text{Menu}_M\}$ 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 10 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 $(p, t) \in \text{Menu}_M$ and its corresponding region $R = \{x : p = P(x)\}$, 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:

$$\hat{B} \in \text{int}(R) \implies \hat{A} \in \text{int}(R); \quad (4)$$
$$\hat{A} \in \text{int}(R) \implies \hat{B} \in \text{cl}(R). \quad (5)$$

This is because, from Lemma 10, 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 10, 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) 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 10 that, almost surely, it holds componentwise

$$\mathbb{E}[\hat{B}|\hat{A}] \geq \hat{A}. \quad (6)$$

The above imply that, almost surely:

$$p_i > 0 \implies \mathbb{E}[\hat{B}|\hat{A}] \cdot \mathbb{I}_{\hat{A} \in \text{int}(R)} = \hat{A}_i \cdot \mathbb{I}_{\hat{A} \in \text{int}(R)} \quad (7)$$

47
and, hence,
\[ \gamma_1 \text{int}(R) \preceq_{\text{cvx}(\tilde{v})} \gamma_2 \text{int}(R) + \xi_R, \]

17It is easy to verify that the guarantees of the lemma remain the same except the third guarantee changes to "componentwise smaller than."

18Notice that the partial derivatives are non-positive.
where similarly to our derivation above \( \xi'_R \) is the non-negative measure corresponding to \( \hat{B}' \).

\[ \mathbb{I}_{A' \in \text{int}(R) \land \hat{B}' \notin \text{int}(R)} \]

We now combine the results of Steps (i) and (ii) to finish the proof. Combining (9) and (13), we get that:

\[ \mu_{+|\text{int}(R)} \preceq_{\text{cvx}(\tilde{\nu})} \mu_{-|\text{int}(R)} + \xi_R + \xi'_R. \]  

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 11, 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 \( \mu_{+}, \xi_R \) and \( \xi'_R \) are non-negative. Therefore, we can rewrite the property (14) as:

\[ \mu_{+|R} \preceq_{\text{cvx}(\tilde{\nu})} \mu_{-|R}. \]

C Missing Proofs of Section 6 - Theorem 5

In this appendix we complete the proof of Theorem 5.

**Proof of Lemma 2:** 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 \( \varphi \) is a bijection. As \( x_n \) ranges from 0 to 1, 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 \( i = 2, \ldots, n-1 \). (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 \( y_1 \geq y_2 \geq \cdots \geq y_n = 0 \). Thus, \( \varphi \) 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 \( \varphi \) 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 \( \varphi' : W \to W \) scales volumes by \( \rho \), where \( W \subset \mathbb{R}^{n-1} \) is the set \( \{ w : 1 \geq w_1 \geq \cdots \geq w_{n-1} \geq 0 \} \) and \( \varphi'(w) \) drops the last (constant) coordinate of \( \varphi(1, w_1, \ldots, w_{n-1}) \) and then (for notational convenience) permutes the first coordinate to the end. That is,

\[
\varphi'(w_1, \ldots, w_{n-1}) = \left( \frac{w_1 - w_{n-1}}{1 - w_{n-1}} z(w_{n-1}), \ldots, \frac{w_{n-2} - w_{n-1}}{1 - w_{n-1}} z(w_{n-1}), z(w_{n-1}) \right)
\]

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

We now analyze the determinant of the Jacobian matrix \( J \) of \( \varphi' \). 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

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

\[
= \left( \frac{z(w_{n-1})}{1 - w_{n-1}} \right)^{n-2} \cdot \frac{-1}{n-1} \left( z(w_{n-1})^{-2} \cdot \rho \cdot (n-1)(1 - w_{n-1})^{n-2} \right) = -\rho
\]

as desired.

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

\( \square \)

**Proof of Theorem 5:** We now complete the proof of Theorem 5. Fix the dimension \( n \). 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 \( (c, c, \ldots, c) \).
- 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 \( \mu^c \) on \( [0,1]^n \) which has mass +1 at the origin, mass of \(-c\) on each each surface \( x_i = 0 \), et cetera. It is important to notice that the mass that \( \mu \) 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 \( \mu_-^{c_0}(Z(h')) \) increases with \( c \) (and becomes arbitrarily large as \( c \) becomes large), there must exist a \( c^0 > c_0 \) such that \( \mu_-^{c'_0}(Z(h')) = 1 \), and thus \( \mu_-^{c'_0}(Z(h')) = 0 \). We can therefore pick a decreasing function \( p^* : \mathbb{R}_{\geq 0} \to (0,1) \) such that, for all sufficiently large...
c, \( \mu^c(Z(p^*(c))) = 0 \). 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 \( p^*(c) \to 0 \) as \( c \to \infty \).

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 \( \mu^c|_{Z_c} \leq cvx \mu^c|_{Z_c} \).

To apply Theorem 3, it remains to show that, for sufficiently large \( c \), \( \mu_+^c|_{W_c} \leq cvx(\overline{\mu}^-) \mu_-^c|_{W_c} \). To prove this, we partition \( W_c \) into \( 2(n!+1) \) disjoint\(^{20} \) regions, \( P_0, P_{\sigma_1}, \ldots, P_{\sigma_n} \) and \( N_0, N_{\sigma_1}, \ldots, N_{\sigma_n} \), where \( \sigma_j \) is a permutation of \( 1, \ldots, n \). 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 \( \mu_+^c|_{P_j} \leq cvx(\overline{\mu}^-) \mu_-^c|_{W_c} \) for all \( j \), thereby proving \( \mu_+^c|_{W_c} \leq cvx(\overline{\mu}^-) \mu_-^c|_{W_c} \).

For every permutation \( \sigma \) of \( 1, \ldots, n \), define:

\[
P_\sigma = \left\{ x : 1 = x_{\sigma(1)} \geq x_{\sigma(2)} \geq \cdots \geq x_{\sigma(n)} \geq 0 \text{ and } x_{\sigma(n)} \leq 1 - \left(\frac{1}{c+1}\right)^{1/(n-1)} \right\}
\]

\[
N_\sigma = \{ y : 1 \geq y_{\sigma(1)} \geq \cdots \geq y_{\sigma(n-1)} \geq y_{\sigma(n)} = 0 \}
\]

Denote by \( \rho \triangleq (c+1)/c \) the ratio between the surface densities of \( \mu_+^c \) and \( \mu_-^c \) on \( P_\sigma \) and \( N_\sigma \), respectively, and let \( \varphi_\sigma : P_\sigma \to N_\sigma \) be the bijection given by Lemma 2. By construction, \( \mu_+^c(S) = \mu_-^c(\varphi_\sigma(S)) \) for all measurable \( S \subseteq P_\sigma \).

Denote \( N_\sigma = N_\sigma' \setminus Z_c \) and \( P_\sigma \triangleq \varphi^{-1}(N_\sigma) \). By construction, \( \varphi \) is a bijection between \( P_\sigma \) and \( N_\sigma \), preserving the respective the measures \( \mu_+^c \) and \( \mu_-^c \), such that for all \( x \in P_\sigma \), \( x \) is componentwise at least as large as \( \varphi(x) \). Therefore, by Strassen’s theorem, \( \mu_+^c|_{P_\sigma} \leq cvx(\overline{\mu}^-) \mu_-^c|_{N_\sigma} \). 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 \( N_\sigma \), and \( N_0 \) consists of all points on which \( \mu_-^c \) is nontrivial which have not yet been matched.\(^{21} \) It therefore remains only to show that \( \mu_+^c|_{P_0} \leq cvx(\overline{\mu}^-) \mu_-^c|_{N_0} \). 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 \( P_\sigma' \).

\(^{19} \)Our 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.

\(^{20} \)For notational simplicity, our regions overlap slightly, although the overlap always has zero mass under both \( \mu_+^c \) and \( \mu_-^c \).

\(^{21} \)All other points on which \( \mu_-^c \) is nontrivial have been matched either to the origin (if the point lies in \( Z_c \)), or to some point in \( P_\sigma \) (if the point lies in \( N_\sigma' \setminus Z_c \)).
• For large $c$, every $x$ with $x_i = 1$ but some $x_j \leq 3/4$ is in some $P'_\sigma$.

• 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 2, we have $\sigma(n) \geq 1 - \left(\frac{p^*(c)^{n-1} + c - 1}{\rho} \right)^{(n-1)}$. As $c$ gets large, $\rho \to 1$ and $p^*(c) \to 0$. 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 $P_\sigma$, 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 6, it suffices to show that for sufficiently large $c$, $Y \geq E[X]$ almost surely.\footnote{In general, to prove second order dominance we might need to nontrivially couple $X$ and $Y$. In this case, however, choosing independent random variables suffices.} Since $\mu^c_\| |_{P_0}$ is supported on $P_0$, we need only show that all coordinates of $E[X]$ are less than $3/4$. We recall that $\mu^c_\|$ 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

$$\frac{\mu^c_\|(Z_c \cap (0,1)^n)}{\mu^c_\|(0,1)^n)} \to 0$$

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

$$\mu^c_\| |_{P_0} \preceq_{\text{cex}(-1)} \mu^c_\| |_{N_0}$$

\[\square\]

D Supplementary Material for Section 7

Proof of Claim 1: 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, the point $(z_x + z_y)/2$ is in...
Thus
\[
u\frac{Z(x+y)}{2} \leq \|\frac{x+y-z_x+zy}{2}\|_1 \leq \|\frac{x-z_x}{2}\|_1 + \|\frac{y-zy}{2}\|_1 = \frac{u_z(x) + u_z(y)}{2}
\]
and therefore \(u_z\) is convex.

Lastly, we verify that \(u_z\) has Lipschitz constant at most 1. Indeed,
\[
u_z(x) - u_z(y) \leq \|x - z_y\|_1 - \|y - z_y\|_1 \leq \|x - y\|_1.
\]
\(\square\)

E Supplementary Material for Section 7

E.1 Verifying Stochastic Dominance - Proof of Lemma 3

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 \(\alpha, \beta\) be positive finite Radon measures on \(\mathbb{R}^n_\geq\) with \(\alpha(\mathbb{R}^n_\geq) = \beta(\mathbb{R}^n_\geq)\). A necessary and sufficient condition for \(\alpha \succeq_1 \beta\) is that for all increasing \(^{23}\) measurable sets \(A\), \(\alpha(A) \geq \beta(A)\).

**Proof of Claim 5:** Without loss of generality assume that \(\alpha(\mathbb{R}^n_\geq) = \beta(\mathbb{R}^n_\geq) = 1\).

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, \(\alpha\) does not stochastically dominate \(\beta\). 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 \(\{r_1, \ldots, r_m\}\) be the set of all values taken by \(\tilde{f}\), where \(r_1 > r_2 > \cdots > r_m\).

We notice that, for any \(s \in \{1, \ldots, m\}\), the set \(A_s = \{z : \tilde{f}(z) \geq r_s\}\) is increasing and measurable. Therefore, we may write
\[
\tilde{f} = \sum_{s=1}^{m} (r_s - r_{s-1}) I_s
\]

\(^{23}\)An increasing set \(A \subset \mathbb{R}^n_\geq\) satisfies the property that for all \(a, b \in \mathbb{R}^n_\geq\) such that \(a\) is component-wise greater than or equal to \(b\), if \(b \in A\) then \(a \in A\) as well.
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 $\int \tilde{f}d\beta > \int \tilde{f}d\alpha$. \hfill $\square$

Due to Claim 5, to verify that a measure $\alpha$ stochastically dominates $\beta$ 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 $z \in \mathbb{R}^n_{\geq 0}$, 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 $\mathbb{R}^n_{\geq 0}$ 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 $\hat{c} = (c_1, \ldots, c_n) \in S$. 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. \hfill $\square$

54
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 \( g, h : \mathbb{R}^n_{\geq 0} \to \mathbb{R}_{\geq 0} \) be bounded integrable functions such that \( \int_{\mathbb{R}^n_{\geq 0}} g(x)dx \) and \( \int_{\mathbb{R}^n_{\geq 0}} h(x)dx \) are finite. Suppose that, for all finite collections \( Z \) of points in \( \mathbb{R}^n_{\geq 0} \), 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 \( A \subseteq \mathbb{R}^n_{\geq 0} \),

\[
\int_A g(x)dx \geq \int_A h(x)dx.
\]

**Proof of Lemma 13:** Let \( A \) be an increasing set. We clearly have \( A = \bigcup z \in A B_z \). For any point \( z \in \mathbb{R}^n_{\geq 0} \), denote by \( z^{n,k} \) the point in \( \mathbb{R}^n_{\geq 0} \) 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 \( A \):

\[
A^l_k \triangleq \bigcup_{z \in A \cap Q_k} B_z; \quad A^u_k \triangleq \bigcup_{z \in A \cap Q_k} B_{z^{n,k}}.
\]

It is clear that both \( A^l_k \) and \( A^u_k \) 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 \( A^l_k \subseteq A \subseteq A^u_k \).

We now will bound

\[
\int_{A^u_k} g(x)dx - \int_{A^l_k} g(x)dx.
\]

Let

\[
W_k = \{ z \in \mathbb{R}^n_{\geq 0} : z_i > k \text{ for some } i \}; \quad W^c_k = \{ z \in \mathbb{R}^n_{\geq 0} : z_i \leq k \text{ for all } i \}.
\]

The set \( W^c_k \) 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

\[
\int_{A^u_k \cap W_k} gdx - \int_{A^l_k \cap W_k} gdx \leq \int_{W_k} gdx.
\]

Furthermore, since \( \lim_{k \to \infty} \int_{W_k} gdx = \int_{\mathbb{R}^n_{\geq 0}} gdx \), we know that \( \lim_{k \to \infty} \int_{W_k} gdx = 0 \). Therefore,

\[
\lim_{k \to \infty} \left( \int_{A^u_k \cap W_k} gdx - \int_{A^l_k \cap W_k} gdx \right) = 0.
\]
Next, we bound
\[ \int_{A_k^u \cap W_k^c} gdx - \int_{A_k^l \cap W_k^c} gdx \leq |g|_{\sup} \left( V(A_k^u \cap W_k^c) - V(A_k^l \cap W_k^c) \right) \]
where $|g|_{\sup} < \infty$ is the supremum of $g$, and $V(\cdot)$ denotes the Lebesgue measure.

For each $m \in \{1, \ldots, n+1\}$ and $z \in \mathbb{R}_{\geq 0}^n$, we define the point $z^{m,k}$ by:
\[
z^{m,k}_i = \begin{cases} 
\max\{0, z_i - 2^{-k}\} & \text{if } i < m \\
z_i & \text{otherwise}
\end{cases}
\]
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,
\[
V(A_k^u \cap W_k^c) - V(A_k^l \cap W_k^c) = \sum_{m=1}^n \left( V(A_k^{m+1} \cap W_k^c) - V(A_k^m \cap W_k^c) \right).
\]

We notice that, for any point $(z_1, z_2, \ldots, z_{m-1}, z_{m+1}, \ldots, z_n) \in [0,k]^{n-1}$, there is an interval $I$ of length at most $2^{-k}$ such that
\[
(z_1, z_2, \ldots, z_{m-1}, w, z_{m-2}, \ldots, z_n) \in (A_k^{m+1} \setminus A_k^m) \cap W_k^c
\]
if and only if $w \in I$. Therefore,
\[
V(A_k^{m+1} \cap W_k^c) - V(A_k^m \cap W_k^c) \leq \int_0^k \cdots \int_0^k 2^{-k} dz_1 \cdots dz_{m-1} dz_{m+1} \cdots dz_n = 2^{-k} k^{n-1}.
\]

We thus have the bound
\[
|g|_{\sup} \left( V(A_k^u \cap W_k^c) - V(A_k^l \cap W_k^c) \right) \leq |g|_{\sup} \sum_{m=1}^n 2^{-k} k^{n-1} = n |g|_{\sup} 2^{-k} k^{n-1}
\]
and therefore
\[
\int_{A_k^u} gdx - \int_{A_k^l} gdx = \int_{A_k^u \cap W_k^c} gdx - \int_{A_k^l \cap W_k^c} gdx + \int_{A_k^u \cap W_k} gdx - \int_{A_k^l \cap W_k} gdx \\
\leq \left( \int_{A_k^u \cap W_k^c} gdx - \int_{A_k^l \cap W_k} gdx \right) + n |g|_{\sup} 2^{-k} k^{n-1}.
\]
In particular, we have
\[
\lim_{k \to \infty} \left( \int_{A_k^u} g \, dx - \int_{A_k^l} g \, dx \right) = 0.
\]
Since \( \int_{A_k^u} g \, dx \geq \int_A g \, dx \geq \int_{A_k^l} g \, dx \), we have
\[
\lim_{k \to \infty} \int_{A_k^u} g \, dx = \int_A g \, dx = \lim_{k \to \infty} \int_{A_k^l} g \, dx.
\]
Similarly, we have
\[
\int_A h \, dx = \lim_{k \to \infty} \int_{A_k^l} h \, dx
\]
and thus
\[
\int_A (g - h) \, dx = \lim_{k \to \infty} \left( \int_{A_k^u} g \, dx - \int_{A_k^l} h \, dx \right).
\]
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 \( \int_{A_k^u} g \, dx - \int_{A_k^l} h \, dx \geq 0 \) for all \( k \). Therefore \( \int_A (g - h) \, dx \geq 0 \), as desired.

We are now ready to prove Lemma 3.

**Proof of Lemma 3:**

We begin by defining, for any \( a \) and \( b \) with \( p_1 \leq a \leq b \leq q_1 \), the function \( \zeta^b_a : [p_2, q_2] \to \mathbb{R} \) by
\[
\zeta^b_a(w_2) \equiv \int_a^b (g(z_1, w_2) - h(z_1, w_2)) \, dz_1.
\]
This function \( \zeta^b_a(w_2) \) 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 \( \zeta^b_a(w_2) \leq 0 \).

**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
\[
\zeta^{\tilde{z}_1}_a(w_2) \leq \zeta^b_a(w_2) \leq \zeta^{q_1}_a(w_2).
\]
We notice, however, that \( \zeta^{q_1}_a(w_2) \leq 0 \) by assumption, and thus the claim is proven.

We now claim the following:
Claim 7. Suppose that \( \zeta^b_a(w^*_2) > 0 \) for some \( w^*_2 \in [c_2, q_2) \). Then \( \zeta^b_a(w_2) \geq 0 \) for all \( w_2 \in [w^*_2, q_2) \).

Proof of Claim 7: Given that \( \zeta^b_a(w^*_2) > 0 \), 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

\[
\zeta^b_a(w_2) = \int_a^b (g(z_1, w_2) - h(z_1, w_2))dz_1 = \int_a^b (\alpha(z_1) \cdot \beta(w_2) \cdot \eta(z_1, w_2))dz_1.
\]

Similarly, \( (c, w^*_2) \notin R \) for any \( c \geq a \), so

\[
\zeta^b_a(w^*_2) = \int_a^b (\alpha(z_1) \cdot \beta(w^*_2) \cdot \eta(z_1, w^*_2))dz_1.
\]

Note that, since \( \zeta^b_a(w^*_2) > 0 \), we have \( \beta(w^*_2) > 0 \). Thus, since \( \eta \) is increasing,

\[
\zeta^b_a(w_2) \geq \int_a^b (\alpha(z_1) \cdot \beta(w_2) \cdot \eta(z_1, w^*_2))dz_1 = \frac{\beta(w_2)}{\beta(w^*_2)} \zeta^b_a(w^*_2) \geq 0,
\]

as desired. \( \square \)

We extend \( g \) and \( h \) to all of \( \mathbb{R}^2_{>0} \) by setting them to be 0 outside of \( C \). By Claim 13, to prove that \( g \geq 1 \) \( h \) it suffices to prove that \( \int_A gdx \geq \int_A hdxdy \) for all sets \( A \) which are the union of finitely many bases. Since \( g \) and \( h \) are 0 outside of \( C \), it suffices to consider only bases \( B_{z'} \) where \( z' \in C \), since otherwise we can either remove the base (if it is disjoint from \( C \)) or can increase the coordinates of \( z' \) moving it to \( C \) without affecting the value of either integral.

We now complete the proof of Lemma 3 by induction on the number of bases in the union.

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

\[
\int_{B_r}(g-h)dx = \int_{r_1}^{q_2} \int_{r_2}^{q_1} (g-h)dz_1dz_2 = \int_{r_1}^{q_2} \zeta^{q_1}_{r_1}(z_2)dz_2.
\]

By Claim 7, we know that either \( \zeta^{q_1}_{r_1}(z_2) \geq 0 \) for all \( z_2 \geq r_2 \), or \( \zeta^{q_1}_{r_1}(z_2) \leq 0 \) 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

\[
\int_{r_2}^{q_2} \zeta^{q_1}_{r_1}(z_2)dz_2 \geq \int_{p_2}^{q_2} \zeta^{q_1}_{r_1}(z_2)dz_2 = \int_{p_2}^{q_2} \int_{r_1}^{q_1} (g-h)dz_1dz_2
\]

\[
= \int_{r_1}^{q_1} \int_{p_2}^{q_2} (g-h)dz_2dz_1.
\]

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 \( z_1 \)

58
between \(p_1\) and \(r_1\). We assume therefore that we are in the second case, and thus
\[
\int_{p_1}^{q_1} \int_{p_2}^{q_2} (g - h)dz_2dz_1 \geq \int_{p_1}^{q_1} \int_{p_2}^{q_2} (g - h)dz_2dz_1 = \int_C (g - h)dxdy,
\]
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
\[
p_1 \leq z_{1}^{(k+1)} < z_{1}^{(k)} < z_{1}^{(k-1)} < \cdots < z_{1}^{(1)}
\]
\[
p_2 \leq z_{2}^{(1)} < z_{2}^{(2)} < z_{2}^{(3)} < \cdots < z_{2}^{(k+1)}.
\]

Figure 8: 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.

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

**Case 1:** \(\xi_{z_{1}^{(k+1)}}(w_2) \leq 0\) for all \(p_2 \leq w_2 \leq z_{2}^{(k+1)}\).
In this case, we see that
\[ \int_{z_2^{(k+1)}} z_2^{(k)} (f - g) dz_1 dz_2 = \int_{z_2^{(k+1)}} z_2^{(k)} \zeta_2^{(k+1)}(w) dw \leq 0. \]

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
\[
(f - g)(A) \geq (f - g)(A) \n + (f - g) \left( \left\{ z : z_1^{(k+1)} \leq z_1^{(k)} \text{ and } z_2^{(k)} \leq z_2^{(k+1)} \right\} \right) 
\]
\[
= (f - g) \left( \bigcup_{i=1}^k B_{z_i^{(k)}} \cup B_{z_1^{(k+1)}, z_2^{(k)}} \right) 
\]
\[
= (f - g) \left( \bigcup_{i=1}^{k-1} B_{z_i^{(k)}} \cup B_{z_1^{(k+1)}, z_2^{(k)}} \right) 
\]
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 \(\geq 0\).

**Case 2:** \(\zeta_2^{(k+1)}(w_2) \geq 0\) for all \(w_2 \geq z_2^{(k+1)}\).

In this case, we have
\[ \int_{z_2^{(k+1)}} z_2^{(k)} (f - g) dz_1 dz_2 = \int_{z_2^{(k+1)}} z_2^{(k)} \zeta_2^{(k+1)}(w) dw \geq 0. \]

Therefore, it follows that
\[ (f - g)(A) = (f - g) \left( \bigcup_{i=1}^k B_{z_i^{(k)}} \right) \n + (f - g) \left( \left\{ z : z_1^{(k+1)} \leq z_1^{(k)} \text{ and } z_2^{(k+1)} \leq z_2^{(k)} \right\} \right) 
\]
\[ \geq (f - g) \left( \bigcup_{i=1}^k B_{z_i^{(k)}} \right) \geq 0, \]
where the final inequality follows from the inductive hypothesis.

\[\square\]

**E.2 Verifying Stochastic Dominance in Example 3**

We sketch the application of Lemma 3 for verifying that \(\mu_+|_W \succeq \mu_-|_W\) in Example 3. We set \(C = [x_{\text{crit}}, 1] \times [y_{\text{crit}}, 1]\) and \(R = Z \cap C\), so that \(W = C \setminus R\). We let \(g\) and \(h\) being the positive
and negative parts of the density function of $\mu|_W$, respectively, so that the density of $\mu|_W$ is given by $g - h$. Since $Z$ lies below both curves $S_{\text{top}}$ and $S_{\text{right}}$, we know that integrating the density of $\mu$ 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 3. In addition, on $W = C \setminus R$, we have

$$g(z_1, z_2) - h(z_1, z_2) = f_1(z_1)f_2(z_2)\left(\frac{2}{1 - z_1} + \frac{3}{1 - z_2} - 12\right)$$

which satisfies the third condition of Lemma 3, as $2/(1 - z_1) + 3/(1 - z_2) - 12$ is increasing. Finally, we verify the first condition of Lemma 3 by integrating $g - h$ over $C$. This integral is equal to $\mu(W) = 0$ and thus all conditions of Lemma 3 are satisfied.

F Extending to Unbounded Distributions

Several results of this paper extend to unbounded type spaces, although such extensions impose additional technical difficulties. Here we briefly discuss how some of our results generalize.

We can often obtain a “transformed measure” (analogous to Theorem 1 even when type spaces are unbounded) using integration by parts. We wish 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 $\lim_{z_i \to \infty} f_i(z_i)z_i^2 \to 0$, 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 $\mu$. It is helpful therefore to consider only measures $\mu$ such that $\int \|x\|_1 d|\mu| < \infty$, to ensure that $\int ud\mu$ 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 $\mu$, 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 $\mu$ 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.