Strong Duality for a Multiple-Good Monopolist

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

September 16, 2014

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 that takes the form of an optimal transportation map between measures. Our framework improves previous partial results, by establishing a *strong* duality theorem between optimal mechanism design and optimal transportation, and is enabled by an extension of the Monge-Kantorovich duality that accommodates convexity constraints. 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 type distribution. This result strengthens several results in the literature, where only sufficient conditions have been provided. As a corollary of our tight characterization of 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]. Finally, we provide a partial characterization of general two-item mechanisms, and use our characterization to obtain optimal mechanisms in several settings, including an example with two independently distributed items where a continuum of lotteries is necessary for revenue maximization.

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

1 Introduction

We study the problem of revenue maximization for a multiple-good monopolist. Given n heterogenous goods and a probability distribution f over $\mathbb{R}^n_{\geq 0}$, we wish to design a mechanism that optimizes the monopolist's revenue against an additive (linear) bidder 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

*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.

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). Importantly, 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, which has identified optimal multi-item mechanisms only in special cases. By providing a strong duality framework, not only do we strengthen the weak duality framework in our prior work [DDT13], but also encompass all other duality-based approaches to the problem [RC98, MV06, GK14]. 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 Lipschitz continuous functions u, representing the utility of the bidder, as in [Roc87]. The objective function of this maximization problem can be written as the expectation of u with respect to a signed measure μ over the type space of the bidder, obtained by an application of the divergence theorem (whose use is standard in this context). Our formulation is summarized in Theorem 1, while Section 2.2 illustrates our formulation in basic settings with independent and with correlated items.

- In Theorem 2, we formulate a dual in the form of an optimal transport problem, and we establish strong duality between the two problems. Roughly speaking, our dual formulation is given the signed measure μ (from Theorem 1) and solves the following minimization problem: (i) first, it is allowed to massage μ into a measure μ' that dominates μ with respect to a particular "convex dominance" stochastic order; (ii) second, it is supposed to find a coupling of the positive part μ'_+ of μ' with its negative part μ'_-; (iii) if a unit of mass of μ'_+ at x is coupled with a unit of mass of μ'_- 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. See Section 8.

- Using our framework, we can provide shorter proofs of optimality of known mechanisms. Instead of reproving an already-known example, in Section 2.2, 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 4, 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).

Grand-Bundling Optimality: Substantial effort in the literature has been devoted to studying optimality of pricing mechanisms as well as finding conditions for grand-bundling optimality; see, e.g., [MV06] and [DDT13] for sufficient conditions under which the mechanism that only sells the grand bundle is optimal. Our second main result (presented in Theorem 3) obtains *necessary and sufficient* conditions for the optimality of the grand-bundling mechanism:

- Theorem 3 establishes the following: the mechanism that offers the grand bundle at price p is optimal *if and only if* the measure μ (derived from the type distribution f according to Theorem 1) 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 μ_- (the negative part of μ) restricted to Z should convexly dominate μ_+ (the positive part of μ) restricted to Z, and μ_+ restricted to W should "second-order stochastically dominate" μ_- restricted to W. In other words, to check grand-bundling optimality it suffices to check 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.

- We illustrate the power of Theorem 3 with Theorem 4, a result that is interesting on its own right, generalizing 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. See Section 5.

- 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 the grand-bundling 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 other classes of mechanisms.

Partial Characterization of Optimal Two-Item Mechanisms: Several optimal two-item mechanisms exhibit a similar underlying structure. Our final contribution is to provide a partial characterization of general two-item mechanisms, extending the corresponding partial characterization of [DDT13].

- In Theorem 5 we provide sufficient conditions under which the optimal two-item mechanism is characterized by the convex set Z of types that receive no items (and pay nothing to the mechanism). In terms of this set, the utility of every type x is $u(x) = \inf_{z \in Z} \|x - z\|_1$. Our conditions are on the measure μ derived from the type distribution f.

- Using Theorem 5 in Section 7.2.2 we present an example with two items distributed independently according to Beta distributions, where the optimal mechanism offers a continuum of lotteries. This example was presented in a preliminary form in [DDT13], but we complete the proof details by providing the complete proof of Lemma 5 in Appendix E.1.

- Our Beta example sheds important light to the menu-size complexity of auctions [HN13], providing a concrete example with independent items in which the optimal mechanism needs to offer uncountably many randomized bundles. Nevertheless, the mechanism can be specified succinctly using our characterization by just providing the set of types of zero utility.

- As illustrated by our Beta example, the sufficient conditions from Theorem 5 can often be "inverted" to derive a candidate optimal mechanism in terms of a set Z, on which the sufficient conditions of our characterization Theorem 5 can be evaluated. In this manner, our characterization is constructive.

- We provide further example applications of the characterization result, including the complete solution for two independent items distributed according to exponential distributions in Section 7.2.3.

2 Revenue Maximization as an Optimization Program

2.1 Setting up the Optimization Program

Our goal is to find the revenue-optimal mechanism $\mathcal{M}$ for selling n goods to a single additive bidder. An additive bidder has a *type* z specifying his value for each good, where z is an element of a *type space* $Z \subseteq \mathbb{R}^n_{\geq 0}$. While the bidder knows his type with certainty, the mechanism designer knows only the probability distribution on Z from which z is drawn. The type of a bidder is sometimes called his *valuation*.

A mechanism consists of two functions: `(i)` an *allocation function* $\mathcal{P} : Z \to [0,1]^n$ specifying the probabilities, for each possible type declaration of the bidder, that the bidder will be allocated each good, and `(ii)` a *price function* $\mathcal{T} : Z \to \mathbb{R}$ specifying, for each declared type of the bidder, the price that he is charged. When an additive bidder of type z declares himself to be of type $z' \in Z$, he receives net utility $z \cdot \mathcal{P}(z') - \mathcal{T}(z')$.

We restrict our attention to mechanisms that are *incentive compatible*, meaning that the bidder must have adequate incentives to reveal his values for the items truthfully, and *individually rational*, meaning that the bidder has an incentive to participate in the mechanism. The key result of the current section is Theorem 1, where, in a manner similar to that of [MV06] and [DDT13], we reduce the optimal mechanism design problem to an optimization problem over feasible utility functions.

Definition 1. *Mechanism $\mathcal{M} = (\mathcal{P}, \mathcal{T})$ over type space $Z \subseteq \mathbb{R}^n_{\geq 0}$ is* incentive compatible (IC) *if and only if $z \cdot \mathcal{P}(z) - \mathcal{T}(z) \geq z \cdot \mathcal{P}(z') - \mathcal{T}(z')$ for all $z, z' \in Z$.*

Definition 2. *Mechanism $\mathcal{M} = (\mathcal{P}, \mathcal{T})$ over type space $Z \subset \mathbb{R}^n_{\geq 0}$ is* individually rational (IR) *if and only if $z \cdot \mathcal{P}(z) - \mathcal{T}(z) \geq 0$ for all $z \in Z$.*

When a buyer truthfully reports his type to a mechanism $\mathcal{M} = (\mathcal{P}, \mathcal{T})$ (over type space Z), we denote by $u : Z \to \mathbb{R}$ the function that maps the buyer's valuation to the utility he receives by $\mathcal{M}$. It follows by the definitions of $\mathcal{P}$ and $\mathcal{T}$ that $u(z) = z \cdot \mathcal{P}(z) - \mathcal{T}(z)$. While our definitions of individual rationality and incentive compatibility apply to mechanisms over an arbitrary type space $Z \subseteq \mathbb{R}^n_{\geq 0}$, we can always extend any such mechanism to a mechanism over a convex type space $X \supseteq Z$. (See Appendix A.) With a convex type space, it is well-known (see [Roc87], [RC98], and [MV06]), that whether or not a mechanism is IC and IR is can be fully determined by inspecting its utility function.

Lemma 1. *Let $\mathcal{M} = (\mathcal{P}, \mathcal{T})$ be a mechanism defined over a convex type space $X \subseteq \mathbb{R}^n_{\geq 0}$. Then $\mathcal{M}$ is IC and IR if and only if the utility function $u(z) \triangleq z \cdot \mathcal{P}(z) - \mathcal{T}(z)$ is convex, nonnegative, nondecreasing, and 1-Lipschitz with respect to the ℓ_1 norm.*

We clarify that a function u is 1-Lipschitz with respect to the ℓ_1 norm if $u(x) - u(y) \le \|x - y\|_1$ for all $x, y \in X$. Given a function u satisfying the conditions of Lemma 1, we can construct an IC and IR mechanism with utility function u. The following claim, originally from [Roc87], is the converse of Lemma 1.

Claim 1. *Let $X \subseteq \mathbb{R}^n_{\ge 0}$ be convex with nonempty interior and let $u : X \to \mathbb{R}$ be a function satisfying the conditions of Lemma 1. Then we can construct an IC and IR mechanism with utility function u by setting $\mathcal{P}(z) = \nabla u(z)$ and $\mathcal{T}(z) = \nabla u(z) \cdot z - u(z)$ wherever ∇u is defined. On the measure-0 set on which ∇u is not defined, we can use an analogous expression for $\mathcal{P}$ and $\mathcal{T}$ by choosing appropriate values of ∇u from the subgradient of u.*

Wherever ∇u is defined, the components of ∇u specify the corresponding allocation probabilities of the mechanism.[1] In Remark 6 of Appendix A, we show that any u satisfying the conditions of Lemma 1 indeed has the property that ∇u is defined almost everywhere, and thus the mechanism corresponding to a feasible u is essentially unique.

We will formulate the mechanism design problem as an optimization problem over feasible utility functions u. We first define the notation:

Definition 3. *Let X be a convex subset of $\mathbb{R}^n$. Then*

- *$\mathcal{U}(X)$ is the set of all functions $u : X \to \mathbb{R}$ which are continuous, non-decreasing, and convex.*
- *$\mathcal{L}_1(X)$ is the set of all functions $f : X \to \mathbb{R}$ which are 1-Lipschitz with respect to the ℓ_1 norm. That is, $f(x) - f(y) \le \|x - y\|_1$ for all $x, y \in X$.*

In this notation, Lemma 1 states that a mechanism $\mathcal{M}$ is IC and IR if and only if its utility function u satisfies $u \ge 0$ and $u \in \mathcal{U}(X) \cap \mathcal{L}_1(X)$.

In the optimal mechanism design problem, the distribution of bidder types is specified by a probability density function over the type space Z. In order to integrate over Z and to use tools such as integration by parts, we must impose mild technical conditions on Z. We call a type space satisfying these conditions *well-behaved*. Well-behaved spaces are bounded, although we can sometimes circumvent this restriction- see Remark 1 and Appendix G.

We recommend the reader not get bogged down in the other details of the definition of a well-behaved type space at the moment, as except for highly correlated distributions the type space is typically an open box, which is well-behaved.

Definition 4. *A type space $U \subseteq \mathbb{R}^n_{\ge 0}$ is* well-behaved *if it is a Jordan-measurable bounded Lipschitz domain. That is, U is open, bounded, connected,[2] and the boundary ∂U both has Lebesgue measure 0 and is locally the graph of a Lipschitz continuous function.*

The following claim generalizes the approach of [MV06] by writing the revenue maximization problem as an optimization problem over feasible utility functions u. The proof is in Appendix A and applies an appropriate version of the divergence theorem (more precisely, a technical variant of integration by parts) to the expression for expected revenue.[3]

[1] Thus, the 1-Lipschitz condition on u is a consequence of the fact that allocation probabilities are at most 1.

[2] We could define a well-behaved type space to be any finite union of Jordan-measurable Lipschitz domains whose closures are pairwise disjoint. This would cause no additional technical difficulties, but is unnecessary for this paper.

[3] This is where we use the condition that U is a Lipschitz domain. Imposing such a condition on the boundary is necessary. Even the basic "textbook" version of the divergence theorem, for example, requires a piecewise smooth boundary.

Claim 2. *Let $U \subset \mathbb{R}^n_{\geq 0}$ be a well-behaved type space and let $f : U \to \mathbb{R}$ be a (differentiable) probability density function with bounded partial derivatives. Let $X = [0, M]^n \supset U$. Then the supremum revenue of an IC and IR mechanism for goods whose values are distributed according to the joint distribution f is given by*

$$\sup_{\substack{u \in \mathcal{U}(X) \cap \mathcal{L}_1(X) \\ u \geq 0}} \left\{ \int_{\partial U} u(z) f(z)(z \cdot \hat{n}) ds - \int_U u(z)(\nabla f(z) \cdot z + (n+1) f(z)) dz \right\}$$

where $\hat{n}$ denotes the outer unit normal field to the boundary ∂U.

Since ∇f is bounded, the expression for expected revenue in Claim 2 is a bounded linear functional of u. By the Riesz representation theorem, we can define a (signed) Radon[4] measure ν (supported within $U \cup \partial U$) so

$$\int h d\nu \triangleq \int_{\partial U} h(z) f(z)(z \cdot \hat{n}) dz - \int_U h(z)(\nabla f(z) \cdot z + (n+1) f(z)) dz$$

for all continuous bounded $h : \mathbb{R}^n \to \mathbb{R}$. With this notation, the expected revenue of a mechanism with utility function u is given by $\int u d\nu$. We notice that ν is supported within X and, by setting h to be the constant function 1, we observe that $\nu(X) = -1$. Since it will be more convenient later to deal with Radon measures of net zero mass, we make Claim 3, whose proof appears in the appendix. This claim says that we can add an appropriate mass to ν so that its net mass becomes zero without significantly affecting our optimization problem.

Claim 3. *Let $X = [0, M]^n$ and let ν be a Radon measure supported within a set $U \subset X$ such that $\nu(X) = -1$. Pick any $z_0 \in X$ such that z_0 is coordinate-wise less than every point in U, and define the measure $\mu \triangleq \nu + \delta_{z_0}$.*[5] *Then $\mu(X) = 0$ and*

$$\sup_{\substack{u \in \mathcal{U}(X) \cap \mathcal{L}_1(X) \\ u(z) \geq 0}} \int_X u d\nu = \sup_{u \in \mathcal{U}(X) \cap \mathcal{L}_1(X)} \int_X u d\mu.$$

Claim 3 simply allows us to consider measures μ with net zero mass, and also lets us disregard the non-negativity constraint on u.[6] We call such a measure μ a *transformed measure* of f.

Definition 5 (Transformed measure)**.** *Let $U \subset \mathbb{R}^n_{\geq 0}$ be a well-behaved type space, let $f : U \to \mathbb{R}_{\geq 0}$ be a probability density function with bounded partial derivatives, and let $z_0 \in \mathbb{R}^n_{\geq 0}$ be any point which is coordinate-wise less than all points in U. We say that a signed Radon measure μ is a* transformed measure of f *if the relation*

$$\int h d\mu = h(z_0) + \int_{\partial U} h(z) f(z)(z \cdot \hat{n}) dz - \int_U h(z)(\nabla f(z) \cdot z + (n+1) f(z)) dz$$

holds for all continuous bounded functions $h : \mathbb{R}^n \to \mathbb{R}$.

A transformed measure μ is supported within $U \cup \partial U$ and $\mu(U \cup \partial U) = 0$. We provide examples of transformed measures in Section 2.2.

Theorem 1 below forms the basis for our approach to optimal mechanism design, and for the remainder of the paper we focus on problems of the form given in the theorem.

[4] A Radon measure is a locally-finite inner-regular Borel measure. We need to use this terminology to be mathematically precise, but it suffices to think of a Radon measure as simply being "nice." Essentially any measure that the reader may think of is Radon.

[5] That is, μ is ν with an added point mass of $+1$ at z_0.

[6] Since $\mu(X) = 0$, adding a constant to u doesn't change the value of $\int u d\mu$.

Theorem 1. *Let $U \subset \mathbb{R}^n_{\geq 0}$ be a well-behaved type space and let $f : U \to \mathbb{R}_{\geq 0}$ be a probability density function with bounded partial derivatives. Then 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 is equivalent to solving the optimization problem*

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

where $X = [0, M]^n \supset U$ and μ is a transformed measure of f.

Remark 1. While Theorem 1 only applies when U is bounded, we can often obtain an analogous result when U is unbounded, as long as the density function f decays sufficiently quickly (to ensure that appropriate integrals and the supremum revenue are finite.) Indeed, many results from this paper can be generalized to the infinite case. See Appendix G for an informal discussion and examples of such extensions. (Note that Appendix G discusses extensions of results from throughout this paper.)

2.2 Examples

We present two concrete examples of constructing transformed measures and applying Theorem 1.

2.2.1 Independent Uniform Items

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 well-behaved set $U = \prod_i (a_i, b_i)$, and we apply Theorem 1 with $X = \prod_i [0, b_i] \supset U$ and $z_0 = (a_1, \ldots, a_n)$.

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

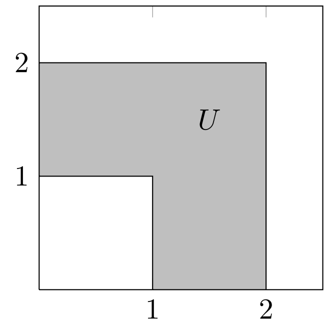
$$\int h d\mu = h(a_1, \ldots, a_n) - \frac{n+1}{v} \int_U h(z) dz + \frac{1}{v} \int_{\partial U} h(z)(z \cdot \hat{n}) ds$$

for all continuous and bounded $h : \mathbb{R}^2 \to \mathbb{R}$. Thus, by Theorem 1, the optimal revenue is $\sup_{u \in \mathcal{U}(X) \cap \mathcal{L}_1(X)} \int_X u d\mu$, where μ is the sum of:

- A point mass of $+1$ at the point $z_0 = (a_1, \ldots, a_n)$.
- A mass of $-(n + 1)$ distributed uniformly throughout the region U.
- A mass of $+\frac{b_i}{b_i - a_i}$ distributed on each surface $\{z \in \partial U : z_i = b_i\}$.
- A mass of $-\frac{a_i}{b_i - a_i}$ distributed on each surface $\{z \in \partial U : z_i = a_i\}$.

2.2.2 Correlated Items with Non-Convex Support

Our framework applies even to joint distributions with non-convex supports. While work such as [RC98] assume that the joint density function f is continuous on a convex support, this restriction is avoidable. Consider two items whose joint distribution is uniform on the set $U \subset \mathbb{R}^2$, the interior of the shaded region in the diagram:



The density function is $f(z) = \frac{1}{3}\mathbb{I}_{z \in U}$, which is continuous on U. We note, however, that f is continuous neither on the convex hull of U nor on the set $X = [0,2]^2$. We can still use Theorem 1 to write the expected revenue achieved by utility function u as $\int u d\mu$, where μ is the sum of:

- A point mass of $+1$ as the origin.
- A mass of -3 uniformly distributed throughout the region U.
- A mass of $+4/3$ uniformly distributed on the line $[0,2] \times \{2\}$ and a mass of $+4/3$ uniformly distributed on the line $\{2\} \times [0,2]$.
- A mass of $-1/3$ uniformly distributed on the line $[0,1] \times \{1\}$ and a mass of $-1/3$ uniformly distributed on the lines $\{1\} \times [0,1]$.

3 The Strong Mechanism Design Duality Theorem

In this section, we present Theorem 2, the main result of this paper. This theorem states that every optimization problem of the form of Theorem 1 has a corresponding dual minimization problem, both of these problems admit optimal solutions, and the optimal values of these problems are equal.

3.1 Measure-Theoretic Preliminaries

We first define the following measure-theoretic notation.

Definition 6. *Let X be a subset of $\mathbb{R}^n$. We define the notation:*

- *$Radon_+(X)$ is the set of all positive Radon measures (locally finite and inner regular Borel measures) supported within X. $Radon(X)$ is the set of all signed Radon measures supported within X.*
- *For any $\gamma \in Radon_+(X \times X)$, the* marginal measures, *denoted $\gamma_1, \gamma_2 \in Radon_+(X)$, are defined by $\gamma_1(S) = \gamma(S, X)$ and $\gamma_2(S) = \gamma(X, S)$ for all measurable $S \subseteq X$. We note that $\gamma_1(X) = \gamma_2(X) = \gamma(X \times X)$.*
- *For any $\alpha, \beta \in Radon_+(X)$ with $\alpha(X) = \beta(X)$, the set $\Gamma(\alpha, \beta)$ is the set of all joint measures $\gamma \in Radon_+(X \times X)$ with respective marginals α and β.*
- *For a (signed) measure $\mu \in Radon(X)$ and a measurable $A \subseteq X$, we define the* restriction of μ to A, *denoted $\mu|_A$, by the property $\mu|_A(S) = \mu(A \cap S)$ for all measurable S.*

- *For a signed measure $\mu \in Radon(X)$, we will denote by $\mu_+, \mu_- \in Radon_+(X)$ the positive and negative parts of μ, respectively. That is, $\mu = \mu_+ - \mu_-$, where μ_+ and μ_- are restrictions of μ to disjoint subsets of X.*[7]

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

Definition 7. *Let X be a bounded convex subset of $\mathbb{R}^n_{\geq 0}$ and $\alpha, \beta \in Radon(X)$. We say that α* convexly dominates *β, denoted $\alpha \succeq_{cvx} \beta$, if for all (non-decreasing, convex) functions $u \in \mathcal{U}(\mathbb{R}^n)$, $\int u d\alpha \geq \int u d\beta$. Similarly, for vector random variables A and B with values in X, we say that $A \succeq_{cvx} B$ if $\mathbb{E}[u(A)] \geq \mathbb{E}[u(B)]$ for all $u \in \mathcal{U}(\mathbb{R}^n)$.*

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, we note that $A \succeq_{cvx} B$ if and only if every *risk-seeking* person with increasing utility "prefers" A to B, while A dominates B in the first order if and only if every person with increasing utility (regardless of risk tolerance) "prefers" A to B. We point out that, since the constant functions ± 1 are in $\mathcal{U}(\mathbb{R}^n)$, the relationship $\alpha \succeq_{cvx} \beta$ for signed measures α and β implies that $\alpha(X) = \beta(X)$.

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 $X = [0, M]^n$ and let $\mu \in Radon(X)$ such that $\mu(X) = 0$. Then*

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

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

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 μ into a new measure μ' with $\mu'(X) = 0$ such that $\mu' \succeq_{cvx} \mu$.
- **Step 2**: Find a joint measure $\gamma \in \Gamma(\mu'_+, \mu'_-)$ such that $\int \|x - y\|_1 d\gamma(x, y)$ is minimized. This is an optimal mass transportation problem where the cost of transporting a unit of mass from a point x to a point y is the ℓ_1 distance $\|x - y\|_1$, and we are asked for the cheapest method of transforming the positive part of μ' into the negative part of μ'.[8] Transportation problems of this form have been studied in the mathematical literature. See [Vil09].

[7] Formally, the decomposition of μ into μ_+ and μ_- is guaranteed to exist uniquely by the Jordan decomposition theorem.

[8] Formally, this interpretation uses the equality

$$\inf_{\substack{\gamma \in Radon_+(X \times X) \\ \gamma_1 - \gamma_2 = \mu'}} \int_{X \times X} \|x - y\|_1 d\gamma = \inf_{\substack{\gamma' \in Radon_+(X \times X) \\ \gamma'_1 = \mu'_+; \quad \gamma'_2 = \mu'_-}} \int_{X \times X} \|x - y\|_1 d\gamma'$$

which follows from Kantorovich-Rubinstein duality, as $\|x - y\|_1$ is a metric.

By Theorem 1 we can rewrite a mechanism design problem as $\sup_{u\in\mathcal{U}(X)\cap\mathcal{L}_1(X)}\int ud\mu$, and by Theorem 2 there always exists an appropriate γ for the dual problem which certifies optimality of the optimal u. We prove Theorem 2 in Section 8. We remark that one direction of the duality theorem is easy. Proving the reverse direction in Section 8 is significantly more challenging, and relies on results such as the Fenchel-Rockafellar duality theorem.

Lemma 2 (Weak Duality)**.** *Let X be a convex subset of $\mathbb{R}^n_{\geq 0}$ and let $\mu \in Radon(X)$ with $\mu(X) = 0$. Then*

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

Proof of Lemma 2: For any feasible u for the left-hand side and feasible γ for the right-hand side, we have

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

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

From the proof of Lemma 2, we note the following "complementary slackness" conditions that a tight pair of optimal solutions must satisfy.

Corollary 1. *Let u^* and γ^* be feasible for their respective problems above. Then $\int u^* d\mu = \int \|x-y\|_1 d\gamma^*$ if and only if both of 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 2 are tight precisely when both conditions hold. □

Remark 2. It is useful to geometrically interpret Corollary 1. The first condition is intricate, but we provide *very* rough intuition. We view $\gamma_1^*-\gamma_2^*$ (denote this by μ') as a "shuffled" μ. That is, (stemming from the $\mu' \succeq_{cvx} \mu$ constraint) we change μ into μ' by repeatedly (1) picking a positive point mass δ_x from μ_+, (2) splitting the point mass into several pieces, and (3) sending the pieces in multiple directions. We require that the center of mass never decreases in any iteration of this process.[9] The constraint $\int u^* d\mu' = \int u^* d\mu$ says that if the center of mass y of the split pieces is strictly larger than x in coordinate i, then $(\nabla u^*)_i = 0$ at x and at y.[10] In addition, whenever a point mass δ_x is split, we can send a piece to a location z only if u^* varies *linearly* along the path from x to z.[11]

The second condition is more straightforward than the first. We view γ^* as a "transport" map between its component measures γ_1^* and γ_2^*. The condition states that if γ^* transports from location

[9] This explanation follows from the interpretation of $\mu' \succeq_{cvx} \mu$ as saying that all risk-seekers "prefer" μ' to μ. This explanation of "splitting" is an oversimplification: μ can also "split" regions of mass (instead of just points of mass), can "merge" negative masses, and so forth.

[10] Technically, we only need that 0 is the i^{th} coordinate of a subgradient in the case that u^* is not differentiable at x or at y. Recall that u^* is convex, and thus $(\nabla u^*)_i = 0$ also at points "in between" x and y.

[11] This condition is a bit more complicated if u^* is not differentiable at x, when we must ensure that u^* varies linearly with the *same slope* between x and all locations to which a piece is sent. That is, the locations to which pieces are sent share the same common subgradient with x.

x to location y, then (since $\nabla u^* \in [0,1]^n$) 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 $\nabla u^*(x)_i = \nabla u^*(y)_i = 1$.[12] That is, the mechanism allocates item i with probability 1 for bidders of type x and for bidders of type y.

By Lemma 2 and Corollary 1, if we can find a "tight pair" of u^* and γ^*, then they are optimal for their respective problems. Theorem 2 shows that this approach always works: for any optimal u^* there always exists a γ^* satisfying the conditions of Corollary 1.

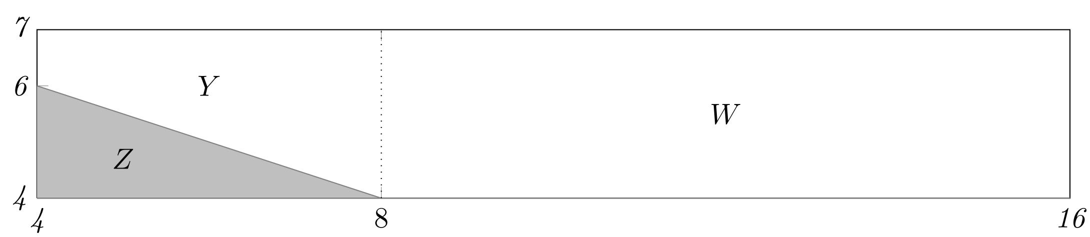
Remark 3. Our duality framework, by achieving strong duality, encompasses all prior duality-based frameworks for optimal mechanism design in our setting [RC98, MV06, DDT13, GK14]. In particular, if we tighten the $\gamma_1 - \gamma_2 \succeq_{cvx} \mu$ constraint in the dual problem to a first-order 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 Example Application of Mechanism Design Duality

We now give an example of using Theorem 2 to prove optimality of a particular mechanism for selling two uniformly distributed independent items. We note that the distributions are not identical, and thus the characterization of [Pav11] does not apply. In addition, the relaxation-based duality framework of [DDT13] (see Remark 3) 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 1. *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 pays a price of 12 and receives both goods.*



This example was constructed for ease of illustration. While our proof, presented next, only verifies the optimality of our proposed mechanism, in Section 6 we develop tools to help us find candidate optimal mechanisms. Drawing inspiration from the work of [Pav11] on iid uniform items and our two-item characterization of Section 6, 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

[12] If u^* is not differentiable at x or at y, we require that u^* has a subgradient with i^{th} coordinate 1 at each of these points.

degenerate one in which only the top edge of the pentagon is non-trivial, resulting in the triangular shape of Z.

Proof of Example 1: It is straightforward to verify that the mechanism described above is IC and IR. All that remains is to prove that the utility function u^* induced by the mechanism is optimal.

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

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

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

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

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

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

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

- Construction of γ^W. We will construct $\gamma^W \in \Gamma(\mu_+|_W, \mu_-|_W)$ such that $x \geq y$ component-wise holds $\gamma^W(x, y)$ almost surely. Geometrically, we view this as "transporting" $\mu_+|_W$ into $\mu_-|_W$ by moving mass downwards and leftwards. Indeed, since both items are allocated with probability 1 in W, being able to transport both downwards and leftwards is in line with our interpretation of the second condition of Corollary 1, as explained in Remark 2.[13]

  We notice that $\mu_+|_W$ consists or mass distributed on the top and right edges of W, while $\mu_-|_W$ consists of mass on the interior and bottom of W. We first match the μ_+ mass on $[8, 16] \times \{7\}$ with the μ_- mass on $[8, 16] \times [\frac{14}{3}, 7]$ by moving mass downwards, then we match the μ_+ mass on $\{16\} \times [4, \frac{14}{3}]$ with the μ_- mass on $[\frac{32}{3}, 16] \times (4, \frac{14}{3}]$ by moving mass to the left, and we finally match the μ_+ mass on $\{16\} \times [\frac{14}{3}, 7]$ with the remaining negative mass arbitrarily. Noticing that $u^*(x) = \|x\|_1 - 12$ for all $x \in W$, it is straightforward to verify the desired properties from Corollary 1.

[13]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.

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

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

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

  The measure α is supported on the line $[4, 8] \times \{7\}$ and consists of a point mass of $\frac{1}{9}$ at $(4, 7)$ followed by allocating mass along the 1-dimensional upper boundary of Y according to a density function which begins negative and increases linearly. It is straightforward to verify that and $\alpha \succeq_{cvx} 0$,[14] 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 γ^Y so that $x_1 = y_1$ and $x_2 \geq y_2$ hold $\gamma^Y(x, y)$ almost surely. The necessary verification, namely that $\mu_+|_Y + \alpha_+$ and $\mu_-|_Y + \alpha_-$ assign the same density to any vertical "strip" in Y, is in Appendix B.

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

□

5 Stochastic Dominance and Grand Bundling Optimality

Identifying conditions under which the optimal mechanism is a simple take-it-or-leave-it offer of the "grand bundle" of all goods has been an important question in the literature; see, e.g., [MV06]. While prior work has identified sufficient conditions for grand bundling optimality, here we do better: Using the framework of Theorem 2, we obtain conditions which are both *necessary* and *sufficient* for grand bundling optimality. In particular, we show in Theorem 3 that grand bundling is optimal if and only if two stochastic dominance relations hold between certain restrictions of the transformed measure μ.

Theorem 2 allows us to certify that a mechanism is optimal by providing a tight γ for the dual problem: The utility function u resulting from a grand bundling mechanism is optimal if and only

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

if a γ exists which satisfies the "complementary slackness" conditions of Corollary 1. Instead of directly providing a tight γ, we obtain conditions on μ for non-constructively proving either that such a γ exists or that no such γ is possible. Such a result is powerful, as verifying properties of μ is typically an easier task than solving an optimization problem over all $\gamma \in Radon_+(X \times X)$.

We stress the importance of this result in that we do not merely derive a sufficient condition to certify optimality, as in [MV06] and [DDT13], but rather we prove that grand bundling optimality is *equivalent* to two measure-theoretic inequalities. The proof of this result is intricate and requires several technical lemmas, which are presented in the appendix. The most important such lemma for our purposes is Lemma 12, and we expect that this lemma can be used to obtain analogous measure-theoretic conditions for optimality in many other classes of natural mechanisms.

Before we state Theorem 3, we define two additional types of stochastic domination.

Definition 8. *Let X be a convex subset of $\mathbb{R}^n_{\geq 0}$ and $\alpha, \beta \in Radon(X)$. We say that α* dominates *β* in the first order*, denoted $\alpha \succeq_1 \beta$, if for all non-decreasing bounded $u : \mathbb{R}^n \to \mathbb{R}$:*

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

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

When X is bounded, we say that α dominates *β* in the second order*, denoted $\alpha \succeq_2 \beta$, if the above inequality holds for all non-decreasing,* concave *functions $u : \mathbb{R}^n \to \mathbb{R}$. We define $A \succeq_2 B$ analogously.*

The definition for second-order dominance is very similar to that of convex dominance presented earlier, except that we test over concave functions instead of convex functions.

We now define two additional properties that a measure μ can satisfy.

Definition 9 (Avoidance of p)**.** *Let μ be a Radon measure on $\mathbb{R}^n$ and let $p > 0$. We say that μ* avoids p *iff $\mu(\{x : \|x\|_1 = p\}) = 0$.*

The property that μ avoids p is incredibly mild in our applications, as discussed in Remark 4. It is included for simplicity of notation and analysis, and we expect that this condition can be removed with additional work.

Definition 10 (Grand Bundling Conditions)**.** *Let $X = [0, M]^n$, let $\mu \in Radon(X)$ with $\mu(X) = 0$, and let $p \in (0, M]$ such that μ avoids p. We say that μ* satisfies the grand bundling conditions with respect to p *iff*

$$\mu|_{\{x \in X : \|x\|_1 \leq p\}} \preceq_{cvx} 0 \quad \text{and} \quad \mu|_{\{x \in X : \|x\|_1 \geq p\}} \succeq_2 0.$$

Theorem 3 (Optimality of Grand Bundling)**.** *Let $U \subset \mathbb{R}^n_{\geq 0}$ be a well-behaved type space, $f : U \to \mathbb{R}$ be a probability density function with bounded partial derivatives, and μ be the transformed measure of f. Let $p > 0$ such that μ avoids p. Then the optimal IC and IR mechanism for a single additive buyer whose values for n goods are distributed according to the joint distribution f is a take-it-or-leave-it offer of the grand bundle at price p* ***if and only if*** *μ satisfies the grand bundling conditions with respect to p.*

Remark 4. The condition that μ avoids p is very mild, and is essentially always satisfied in our mechanism design applications. Given a density function f as in Theorem 1, the transformed measure μ only has "surface density" on the boundary ∂U of U and at a particular point z_0. The avoidance condition will be satisfied unless $\|z_0\|_1 = p$ (which is a degenerate case of grand bundling, and simple analysis yields a condition similar to Theorem 3) or if ∂U intersects $\{x : \|x\|_1 = p\}$ on a region with non-zero $(n-1)$-dimensional surface measure. This only occurs for highly contrived correlated distributions f. We expect that the avoidance condition can be removed entirely with additional work, although such an extension would only be necessary for analyzing very special correlated distributions.

The heart of the proof of Theorem 3 is Lemma 3, a full proof of which is presented in Appendix C.

Lemma 3. *Let $X = [0, M]^n$, let $\mu \in Radon(X)$ with $\mu(X) = 0$, and let $p \in (0, M]$. Define the function $u_p : X \to R_{\geq 0}$ by $u_p(x) = \max\{\|x\|_1 - p, 0\}$ and define the regions $Z, P, W \subset X$ by*

$$Z = \{x \in X : \|x\|_1 \leq p\}\,; \quad P = \{x \in X : \|x\|_1 = p\}\,; \quad W = \{x \in X : \|x\|_1 \geq p\}\,.$$

Suppose that $\mu(P) = 0$. Then the following conditions are equivalent:

1. *There exists $\gamma \in Radon_+(X \times X)$ such that:*
   - $\gamma_1 - \gamma_2 \succeq_{cvx} \mu$
   - $\int u_p d\gamma_1 - \int u_p d\gamma_2 = \int u_p d\mu$, *and*
   - $u_p(x) - u_p(y) = \|x - y\|_1$, *$\gamma(x, y)$-almost surely.*
2. $\mu_-|_Z \succeq_{cvx} \mu_+|_Z$ *and* $\mu_+|_W \succeq_2 \mu_-|_W$.

In particular, if either of the two above conditions are satisfied, then u_p maximizes $\sup_{u \in \mathcal{U}(X) \cap \mathcal{L}_1(X)} \int_X u d\mu$.

The function u_p in Lemma 3 is the utility function corresponding to the grand bundling mechanism with price p, and the conditions on the measure $\gamma \in Radon_+(X \times X)$ are precisely the conditions necessary to certify optimality by Corollary 1. By the strong duality of Theorem 2, grand bundling is optimal if and only if such a γ exists.

One direction of Lemma 3 is easier to prove than the other: given μ satisfying the stochastic domination properties of Condition 2, it is not very difficult (albeit still technical) to construct an appropriate γ. The construction uses the stochastic dominance relations on Z and W to define γ separately on $Z \times Z$ and $W \times W$. The difficulty is in proving that Condition 2 is necessary: given an optimal γ, it is not obvious that the stochastic dominance relation $\gamma_1 - \gamma_2 \succeq_{cvx} \mu$ holds when we restrict γ_1, γ_2, and μ to either of the regions Z or W,[15] and it is conceivable that γ transports mass from one region to the other. We must prove that, whenever a certifying γ exists, it can be appropriately decomposed. The proof requires several technical lemmas, and is presented in Appendix C.

Example of Grand Bundling Optimality

We now present an example of applying our characterization of grand bundling optimality. This result applies to a setting with arbitrarily many items, which is relatively rare in the optimal mechanism design literature.

Theorem 4. *For any integer $n > 0$ there exists a c_0 such that for all $c > c_0$, 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.*

[15] If this stochastic dominance relation held separately in the two regions, it would be possible, with a little work, to deduce Condition 2.

Remark 5. [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 4 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 4 appears in the Appendix C.4.

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

$$A = \left\{ x : 1 = x_1 \geq x_2 \geq \cdots \geq x_n \text{ and } x_n \leq 1 - \left(\frac{\rho - 1}{\rho}\right)^{1/(n-1)} \right\}$$
$$B = \{ y : y_1 \geq \cdots \geq y_n = 0 \}.$$

There exists a continuous bijective map $\varphi : A \to B$ 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 $(\varphi(x))_1 \leq \epsilon$ then $x_n \geq 1 - \left(\frac{\epsilon^{n-1} + \rho - 1}{\rho}\right)^{1/(n-1)}$.*

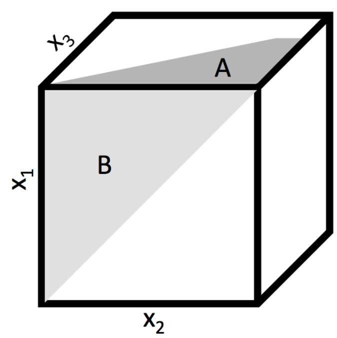


Figure 1: The regions of Lemma 4 for the case $n = 3$.

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

6 A Characterization of Two Item Auctions

In several two-item settings, we find that the optimal mechanisms share a common structure. Our goal in this section is to explain that structure and the geometric intuition behind it.[16]

We first define the notion of a zero set. Every zero set gives rise to a mechanism where the utility of a bidder is equal to the ℓ_1 distance between the bidder's type and the closest point in the zero set. (Bidder types within the zero set receive zero utility, and hence the name.)

[16]The structural result of this section is a strengthening of the two-item result from [DDT13].

Definition 11 (Zero set)**.** *Let $X = [0, M]^2 \subset \mathbb{R}^2_{\geq 0}$. A* zero set Z of X *is a convex, compact, and decreasing*[17] *subset of X with nonempty interior.*

Every zero set Z of X corresponds to a particular mechanism:

Definition 12 (Mechanism of a Zero Set)**.** *A zero set Z of X induces a mechanism whose utility function $u_Z : X \to \mathbb{R}$ is defined by:*

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

Since a zero set Z is closed, for any $x \in X$ there exists a $z \in Z$ such that $u_Z(x) = \|z - x\|_1$.

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

To provide sufficient conditions for u_Z to be optimal, 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 2 is straightforward.

Definition 13 (Outer boundary function of a zero set)**.** *Let $X = [0, M]^2 \subset \mathbb{R}^2$ and let Z be a zero set of X. Denote by $c_1 \in [0, M]$ the point $c_1 = \max\{c : (c, 0) \in Z\}$. We define the* outer boundary function of Z *to be the function $s : [0, c_1] \to [0, M]$ given by*

$$s(y_1) = \max\{y_2 : (y_1, y_2) \in Z\}.$$

Notice that $s(c_1) = 0$ unless Z has a "vertical edge" at c_1.

Definition 14 (Critical value, Canonical partition)**.** *Let Z be a zero set of X with outer boundary function $s : [0, c_1] \to [0, M]$, as in Definition 13. Denote by $a_1, b_1 \in [0, c_1]$ the points such that*

- $0 \geq s'(z_1) > -1$ *for (almost) all* $z_1 \in [0, a_1)$
- $s'(z_1) = -1$ *for all* $z_1 \in (a_1, b_1)$
- $-1 > s'(z_1)$ *for (almost) all* $z_1 \in (b_1, c_1]$.

We call b_1 the critical value of Z. *We define* the canonical partition of X induced by Z *to be the partition of X into $Z \cup \mathcal{A} \cup \mathcal{B} \cup \mathcal{W}$, where*

$$\mathcal{A} = [0, a_1) \times [0, M] \setminus Z; \quad \mathcal{B} = [b_1, M] \times [0, s(b_1)) \setminus Z; \quad \mathcal{W} = X \setminus (Z \cup \mathcal{A} \cup \mathcal{B}),$$

as shown in Figure 2.

Note that the outer boundary function s of a zero set Z is concave and thus is differentiable almost everywhere on $[0, c_1]$ and has non-increasing derivative.

We now restate the utility function u_Z in terms of a canonical partition.

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

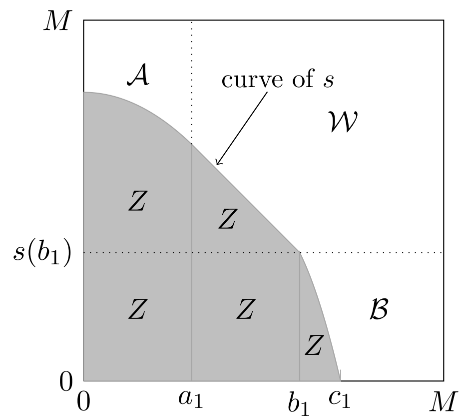


Figure 2: A canonical partition of $[0, M]^2$

Claim 5. *Let Z be a zero set of X with outer boundary function s, and let $Z \cup \mathcal{A} \cup \mathcal{B} \cup \mathcal{W}$ be a canonical partition. Then for all $(v_1, v_2) \in X$:*

$$u_Z(v_1, v_2) = \begin{cases} 0 & \text{if } (v_1, v_2) \in Z \\ v_2 - s(v_1) & \text{if } (v_1, v_2) \in \mathcal{A} \\ v_1 - s^{-1}(v_2) & \text{if } (v_1, v_2) \in \mathcal{B} \\ v_1 + v_2 - (a_1 + s(a_1)) & \text{if } (v_1, v_2) \in \mathcal{W}. \end{cases}$$

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

Pick any $v = (v_1, v_2) \in \mathcal{A}$. We will show that the closest $z \in Z$ is the point $z^* = (v_1, s(v_1))$. 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. Furthermore, we know that z' lies on the boundary of Z, and thus $z' = (v_1 - \delta, s(v_1 - \delta))$ for some $\delta \geq 0$. We may assume that $\delta \leq v_2 - s(v_1)$.

Since $s'(\cdot) \geq -1$ in the range $[0, v_1]$, we know that $s(v_1 - \delta) \leq s(v_1) + \delta \leq v_2$. Therefore

$$\|v - z'\|_1 = \delta + |v_2 - s(v_1 - \delta)| \geq \delta + (v_2 - s(v_1) - \delta) = v_2 - s(v_1).$$

□

We now describe sufficient conditions under which u_Z is optimal.

Definition 15 (Well-formed canonical partition)**.** *Let $Z \cup \mathcal{A} \cup \mathcal{B} \cup \mathcal{W}$ be a canonical partition of $X = [0, M]^2$ induced by zero set Z and let μ be a signed Radon measure on X such that $\mu(X) = 0$. We say that the canonical partition is* well-formed with respect to μ *if the following conditions are satisfied:*

1. $\mu(Z) = 0$
2. $0 \succeq_{cvx} \mu|_Z$
3. $\mu|_{\mathcal{W}} \succeq_2 0$
4. *For all $v \in X$ and all $\epsilon > 0$, $\mu|_{\mathcal{A}}([v_1, v_1 + \epsilon] \times [v_2, M]) \geq 0$, with equality whenever $v_2 = 0$.*

5. *For all $v \in X$ and all $\epsilon > 0$, $\mu|_{\mathcal{B}}([v_1, M] \times [v_2, v_2 + \epsilon]) \geq 0$, with equality whenever $v_1 = 0$.*

We point out the similarities between a well-formed canonical partition and the characterization of grand bundling optimality of Theorem 3. When Z is the zero-set of a bundling mechanism (so $Z = \{z : \|z\|_1 \leq p\}$), $\mathcal{A}$ and $\mathcal{B}$ are empty and the conditions of a well-formed canonical partition are essentially the same as those of Theorem 3 in the two-item case. We interpret Conditions 4 and 5 as saying that $\mu|_{\mathcal{A}}$ (resp. $\mu|_{\mathcal{B}}$) allows for the positive mass in any vertical (resp. horizontal) "strip" to be matched to the negative mass in the strip by only transporting "downwards" (resp. "leftwards").[18] In practice, when μ is given by a density function, we verify these conditions by analyzing the integral of the density function along appropriate vertical or horizontal lines.

Theorem 5. *Let $U \subset \mathbb{R}^2$ be a well-behaved type space and let f be a probability density function on U with bounded partial derivatives. Let $X = [0, M]^2 \supset U$ and let μ be the transformed measure of f. If there exists a zero set Z inducing a canonical partition $Z \cup \mathcal{A} \cup \mathcal{B} \cup \mathcal{W}$ of X that is well-formed with respect to μ, then the optimal IC and IR mechanism for a single additive buyer whose values for two goods are distributed according to the joint distribution f is the mechanism induced by zero set Z. In particular, the mechanism uses the following allocation and price for a bidder with reported type $(z_1, z_2) \in X$:*

- *if $(z_1, z_2) \in Z$, the bidder receives no goods and is charged 0;*
- *if $(z_1, z_2) \in \mathcal{A}$, the bidder receives item 1 with probability $-s'(z_1)$, item 2 with probability 1, and is charged $s(z_1) - z_1 s'(z_1)$;*
- *if $(z_1, z_2) \in \mathcal{B}$, the bidder receives item 1 with probability 1, item 2 with probability $-1/s'(s^{-1}(z_2))$, and is charged $s^{-1}(z_2) - z_2/s'(s^{-1}(z_2))$;*
- *if $(z_1, z_2) \in \mathcal{W}$, the bidder receives both goods with probability 1 and is charged $b_1 + s(b_1)$;*

where s is the outer boundary function of Z and b_1 is the critical value of Z.

Refer back to Figure 2 to visualize such a mechanism.

Proof of Theorem 5: We will show that u_Z maximizes $\sup_{u \in \mathcal{U}(X) \cap \mathcal{L}_1(X)} \int_X u d\mu$. By Corollary 1, it suffices to provide a $\gamma \in Radon_+(X \times X)$ such that $\gamma_1 - \gamma_2 \succeq_{cvx} \mu$, $\int u_Z d(\gamma_1 - \gamma_2) = \int u_Z d\mu$, and $u_Z(x) - u_Z(y) = \|x - y\|_1$ holds γ-almost surely. The γ we construct will never transport mass between regions. That is, $\gamma = \gamma_Z + \gamma_{\mathcal{W}} + \gamma_{\mathcal{A}} + \gamma_{\mathcal{B}}$ where[19]

- $\gamma_Z = 0$. We notice that $(\gamma_Z)_1 - (\gamma_Z)_2 = 0 \succeq_{cvx} \mu|_Z$.
- $\gamma_{\mathcal{W}}$ is constructed such that $(\gamma_{\mathcal{W}})_1 - (\gamma_{\mathcal{W}})_2 \succeq_{cvx} \mu|_{\mathcal{W}}$ and the component-wise inequality $x \geq y$ holds $\gamma_{\mathcal{W}}(x, y)$ almost surely.[20] As in our proof of Theorem 3, the existence of such a $\gamma_{\mathcal{W}}$ is guaranteed by the techniques of Theorem 7.A.3 and Theorem 4.A.6 of [SS10] applied to second order dominance, combined with Strassen's theorem for first-order stochastic dominance. In particular, the condition $\mu|_{\mathcal{W}} \succeq_2 0$ implies the existence of a signed measure θ on $\mathcal{W}$ such that $\theta(\mathcal{W}) = 0$, $\int \|x\|_1 d\theta = \int \|x\|_1 d\mu|_{\mathcal{W}}$, and $0 \preceq_1 \theta \preceq_2 \mu|_{\mathcal{W}}$, where $\succeq_1$ denotes first-order

[18] Notice that these conditions are different from the standard first-order dominance transport criteria, as while we may transport "downwards" to match $\mu|_{\mathcal{A}}$ we may not transport "leftwards," and vice versa for $\mu|_{\mathcal{B}}$. Intuitively, our condition states that the positive part of $\mu|_A$ dominates its negative part when restricted to arbitrarily thin vertical strips, and an analogous condition holds for $\mu|_B$ for horizontal strips.

[19] We have chosen this notation for simplicity, where $\gamma_Z \in Radon_+(Z \times Z)$, $\gamma_{\mathcal{W}} \in Radon_+(\mathcal{W} \times \mathcal{W})$, and so on.

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

stochastic dominance. Since $\theta_+ \succeq_1 \theta_-$, we use Strassen's theorem on first-order stochastic dominance to construct an appropriate map $\gamma_{\mathcal{W}}$ with respective marginals θ_+ and θ_-.

We have that $(\gamma_{\mathcal{W}})_1 - (\gamma_{\mathcal{W}})_2 = \theta \preceq_2 \mu|_{\mathcal{W}}$. Furthermore, $\int \|x\|_1 d\theta = \int \|x\|_1 d\mu|_{\mathcal{W}}$. It follows from the argument of Lemma 8 in the Appendix (used in the proof of Theorem 3) that $(\gamma_{\mathcal{W}})_1 - (\gamma_{\mathcal{W}})_2 = \theta \succeq_{cvx} \mu|_{\mathcal{W}}$. In addition, since $\mu|_{\mathcal{W}}(\mathcal{W}) = \theta(\mathcal{W}) = 0$ and since $u_Z(x)$ differs from $\|x\|_1$ by a constant on $\mathcal{W}$, we have $\int u_Z d\mu|_{\mathcal{W}} = \int u_Z d\theta$.[21]

- $\gamma_{\mathcal{A}} \in Radon_+(\mathcal{A} \times \mathcal{A})$ will be constructed to have respective marginals $\mu_+|_{\mathcal{A}}$ and $\mu_-|_{\mathcal{A}}$, and so that, $\gamma_{\mathcal{A}}(x, y)$ almost surely, it holds that $x_1 = y_1$ and $x_2 \geq y_2$. Thus, $(\gamma_{\mathcal{A}})_1 - (\gamma_{\mathcal{A}})_2 = \mu|_{\mathcal{A}}$, and $\gamma_{\mathcal{A}}$ sends positive mass "downwards."[22] We claim that such a map can indeed be constructed, by noticing that Property 4 of Definition 15 guarantees that, restricted to any vertical strip inside $\mathcal{A}$, μ_+ first-order stochastically dominates μ_-.[23] Hence, Strassen's theorem guarantees that restricted to that strip μ_+ can be coupled with μ_- so that, with probability 1, mass is only moved downwards.

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

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

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

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

7 Applying the Structural Result

In this section, we provide example applications of Theorem 5. A technical difficulty is verifying the stochastic dominance relation $\mu|_{\mathcal{W}} \succeq_2 0$ required to apply the theorem. In our examples, we actually show $\mu|_{\mathcal{W}} \succeq_1 0$, which is easier to verify, yet still imposes technical difficulties. Thus, Section 7.1 presents a useful tool, Lemma 5, for verifying first-order stochastic dominance. Then, Section 7.2 provides example applications of Theorem 5, where this tool is used.

7.1 Verifying First-Order Stochastic Dominance

The following lemma presents a useful tool for verifying first order stochastic dominance between measures.[24]

Lemma 5. *Let $\mathcal{C} = [p_1, q_1) \times [p_2, q_2) \subseteq \mathbb{R}^2_{\geq 0}$ where q_1 and q_2 are possibly infinite, let R be a decreasing nonempty subset of $\mathcal{C}$, and let $g, h : \mathcal{C} \to \mathbb{R}_{\geq 0}$ be bounded integrable functions which are 0 on R, satisfy $\int_{\mathcal{C}} g(x, y)dxdy < \infty$ and $\int_{\mathcal{C}} h(x, y)dxdy < \infty$, and satisfy the following conditions.*

- $\int_{\mathcal{C}} (g - h)dxdy \geq 0$.

[21]We remark that the construction of θ is the only part of our proof of Theorem 5 where we use the fact that X is bounded.

[22]Once again, the intuition for this construction follows Remark 2.

[23]Indeed, as $\epsilon \to 0$, Property 4 states exactly the one-dimensional equivalent condition for first-order stochastic dominance in terms of cumulative density functions.

[24]The lemma also appeared as Theorem 7.4 of [DDT13] without a proof. We provide a detailed proof in Appendix E.1.

- *For any basis vector $e_i \in \{e_1 \equiv (1,0), e_2 \equiv (0,1)\}$ and any point $z^* \in R$:*

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

- *There exist non-negative functions $\alpha : [p_1, q_1) \to \mathbb{R}_{\ge 0}$ and $\beta : [p_2, q_2) \to \mathbb{R}_{\ge 0}$, and an increasing function $\eta : \mathcal{C} \to \mathbb{R}$ such that*

$$g(z_1, z_2) - h(z_1, z_2) = \alpha(z_1) \cdot \beta(z_2) \cdot \eta(z_1, z_2)$$

  for all $(z_1, z_2) \in \mathcal{C} \setminus R$.

Then $\kappa \succeq_1 \lambda$, where κ and λ are the measures corresponding to the density functions g and h respectively.

This result provides a sufficient condition for a measure to stochastically dominate another in the first order. A complete proof of Lemma 5 is in Appendix E.1 and is an application of Claim 14, 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 5 are satisfied, we can induct on the number of boxes by removing one box at a time. We note that Lemma 5 is applicable even to distributions with unbounded support.

Lemma 5 deals with the scenario where two density functions, g and h, are both nonzero on some set $\mathcal{C} \setminus R$, where R is a decreasing subset of $\mathcal{C}$. This setup is motivated by Figure 2. Recall that, in order to apply Theorem 5, we need to check a second order stochastic dominance condition in region $\mathcal{W}$, namely $\mu|_{\mathcal{W}} \succeq_2 0$. Instead, it suffices to show the first order stochastic dominance $\mu|_{\mathcal{W}} \succeq_1 0$, which we plan to show via Lemma 5 by taking $\mathcal{C} = [a_1, M] \times [s(b_1), M]$, $R = \mathcal{C} \cap Z$, and g, h the densities corresponding to measures $(\mu|_{\mathcal{W}})_+$ and $(\mu|_{\mathcal{W}})_-$. To prove that g dominates h in the first order, Lemma 5 states that it suffices to verify that (1) $g - h$ has an appropriate form; (2) the integral of $g - h$ on $\mathcal{C}$ is *positive*; 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*.

7.2 Examples

We apply Theorem 5 to obtain optimal mechanisms in several two-item settings. In Section 7.2.1, we show how using our framework we can easily verify the optimality of the mechanism of [MV06] for two independent uniform $[0, 1]$ items. In Section 7.2.2, we consider two independent items distributed according to different beta distributions. We find the optimal mechanism, showing that it actually offers an uncountably infinite menu of lotteries. We conclude with Section 7.2.3 where we discuss extensions of Theorem 5 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.

7.2.1 A Warm-Up Example: Two Uniform $[0, 1]$ Items

Using Theorem 5, we provide a short proof of optimality of the mechanism for two i.i.d. uniform $[0, 1]$ items proposed by Manelli and Vincent [MV06], which offers each item separately for a price of $\frac{2}{3}$ and the bundle of both items for a price of $\frac{4-\sqrt{2}}{3}$, henceforth called the MV-mechanism.

Let Z be the set of types that receive no goods and pay 0 to the MV-mechanism. Set Z is illustrated in Figure 3 in gray color. Observe that the MV-mechanism is identical to the mechanism

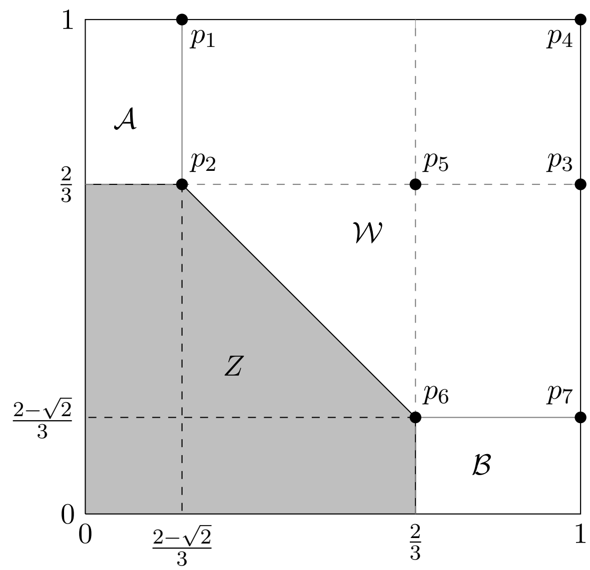


Figure 3: The well-formed canonical partition for two i.i.d. uniform $[0, 1]$ items.

induced by Z according to Definition 12. Now let us show that it is optimal. First we find the canonical partition induced by Z according to Definition 14. The partition is shown in Figure 3 (ignore the dashed lines). To prove optimality of the MV-mechanism, Theorem 5 states that it suffices to show that the partition is well-formed with respect to μ, the transformed measure of the type distribution. What is μ? We have already computed μ in Section 2.2.1. It has a point mass of $+1$ at $(0, 0)$, a mass of -3 distributed uniformly over $[0, 1]^2$, a mass of $+1$ distributed uniformly on the top boundary of $[0, 1]^2$, and a mass of $+1$ distributed uniformly on the right boundary.

Now let us check the conditions of Definition 15. Condition 4 is trivial to verify. Indeed, integrating μ over any rectangle that intersects the top boundary of $[0, 1]^2$ and is contained in $\mathcal{A}$ results in a non-negative integral, given that the intercept of the bottom boundary of region $\mathcal{A}$ is $2/3$. A similar argument shows that Condition 5 holds in region $\mathcal{B}$. Condition 1 is also trivial to verify, given that the surface area of region Z is $1/3$, hence there is a total of -1 negative mass spread uniformly over Z, and a point mass of $+1$ at the origin. By the same token $(\mu|_Z)_- \succeq_1 (\mu|_Z)_+$, and hence Condition 2 is also satisfied. It remains to show Condition 3. We will actually show that $(\mu|_\mathcal{W})_+ \succeq_1 (\mu|_\mathcal{W})_-$ by describing an explicit matching between $(\mu|_\mathcal{W})_+$ and $(\mu|_\mathcal{W})_-$ that only moves mass downwards and leftwards. 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 $\mathcal{W}$ has been accounted for, all of $(\mu|_\mathcal{W})_+$ has been matched to all of $(\mu|_\mathcal{W})_-$ and all moves were down and to the left, establishing $(\mu|_\mathcal{W})_+ \succeq_1 (\mu|_\mathcal{W})_-$.

7.2.2 An Optimal Auction with Infinite Menu Size: Two Beta Items

In the previous section, we used Theorem 5 to prove optimality of a given mechanism. What if we wanted to both calculate the optimal mechanism and certify its optimality from scratch? In this section, we calculate the optimal mechanism for two items distributed according to Beta distributions. In doing so we illustrate a general approach for finding closed-form descriptions of optimal mechanisms via the following steps: `(i)` definition of the sets S_{top} and S_{right}, `(ii)` computation of a critical price p^*, `(iii)` definition of a canonical partition in terms of `(i)` and `(ii)`, and `(iv)` application of Theorem 5. Our approach succeeds in pinning down optimal mechanisms in all examples considered in Sections 7.2.1—7.2.3, 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 bidder a menu of uncountably infinitely many lotteries to choose from. Using our approach we can nevertheless efficiently compute and describe the optimal mechanism.

Let us proceed to our example setting. We consider two items whose values are distributed independently according to the following density functions:

$$f_1(z_1) = \frac{1}{B(3,3)} z_1^2(1-z_1)^2; \qquad f_2(z_2) = \frac{1}{B(3,4)} z_2^2(1-z_2)^3$$

where $z_i \in (0,1)$, and $B(\cdot,\cdot)$ is the Beta function, used for normalization.

To find the optimal mechanism for our example setting, we first find the measure μ induced by f. Notice that $-\nabla f(z)\cdot z - 3f(z) = f_1(z_1)f_2(z_2)\left(\frac{2}{1-z_1} + \frac{3}{1-z_2} - 12\right)$, and $f(z)z = \vec{0}$ at $z_i = 0$ or $z_i = 1$; thus the transformed measure μ comprises:

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

$$f_1(z_1)f_2(z_2)\left(\frac{2}{1-z_1} + \frac{3}{1-z_2} - 12\right).$$

Note that the density of μ is positive on $\mathcal{X} = \left\{z \in (0,1)^2 : \frac{2}{1-z_1} + \frac{3}{1-z_2} > 12\right\} \cup \{\vec{0}\}$ and non-positive on $\mathcal{Y} = \left\{z \in [0,1)^2 \setminus \{\vec{0}\} : \frac{2}{1-z_1} + \frac{3}{1-z_2} \le 12\right\}$, and that $\mathcal{Y} \cup \{\vec{0}\}$ is a decreasing set.

Step (i). We define the set $S_{\text{top}} \subset [0,1)^2$ by the rule that $(z_1,z_2) \in S_{\text{top}}$ iff $\int_{z_2}^1 \mu(z_1,t)dt = 0$. That is, starting from any point $z \in S_{\text{top}}$ and integrating the density of μ "upwards" from $t = z_2$ to $t = 1$ yields zero. Since $\mathcal{Y} \cup \{\vec{0}\}$ is a decreasing set, it follows that $S_{\text{top}} \subset \mathcal{Y}$ and that integrating μ upwards starting from any point above S_{top} yields a positive integral. Similarly, we say that $(z_1,z_2) \in S_{\text{right}}$ iff $\int_{z_1}^1 \mu(t,z_2)dt = 0$, noting that $S_{\text{right}} \subset \mathcal{Y}$. S_{top} and S_{right} are shown in Figure 4.

We analytically compute that $(z_1,z_2) \in S_{\text{top}}$ if and only if

$$z_1 = \frac{2(-1-3z_2-6z_2^2+25z_2^3)}{3(-1-3z_2-6z_2^2+20z_2^3)}.$$

Similarly, $(z_1,z_2) \in S_{\text{right}}$ if and only if $z_2 = \frac{2(-2-4z_1-6z_1^2+27z_1^3)}{-7-14z_1-21z_1^2+72z_1^3}$.

In particular, for any $z_1 \in [0, .63718)$ there exists a z_2 such that $(z_1,z_2) \in S_{\text{right}}$, and there does not exist such a z_2 if $z_1 > .63718$. Furthermore, it is straightforward to verify (by computing second derivatives in the appropriate regime) that the region below S_{top} and the region below S_{right} are convex.

Step (ii). We now calculate $p^* = 0.71307$ as the intercept of the 45° line in Figure 4 which causes $\mu(Z) = 0$ for the set $Z \subset [0,1]^2$ lying below S_{top}, S_{right} and the 45° line. We also define the set $L = \left\{z \in [0,1]^2 : z_1 + z_2 = p^*\right\}$, and calculate $L \cap S_{\text{top}} = \{(.16016, .55291)\}$ and $L \cap S_{\text{right}} =$

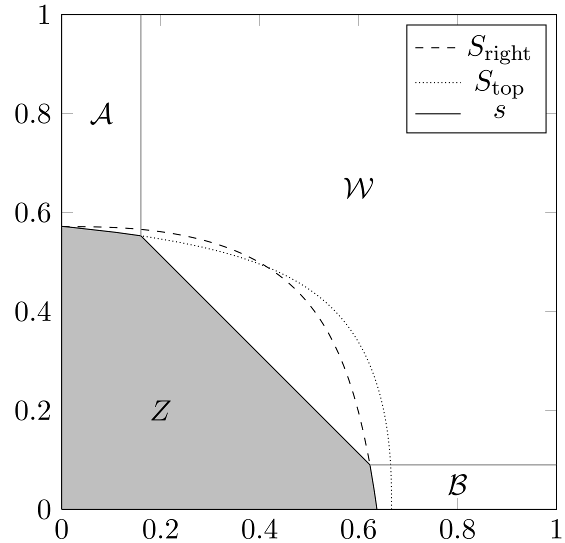


Figure 4: The well-formed canonical partition for $f_1(z_1) = \frac{z_1^2(1-z_1)^2}{B(3,3)}$ and $f_2(z_2) = \frac{z_2^2(1-z_2)^3}{B(3,4)}$.

$\{(.62307, 0.09)\}$. We now define the curve $s : [0, .63718] \to [0, 1]$ by

$$s(z_1) = \begin{cases} z_2 \text{ such that } (z_1, z_2) \in S_{\text{top}} & \text{if } 0 \le z_1 \le .16016 \\ .71307 - z_1 & \text{if } .16016 \le z_1 \le .62307 \\ z_2 \text{ such that } (z_1, z_2) \in S_{\text{right}} & \text{if } .62307 \le z_1 \le .63718. \end{cases}$$

It is straightforward to verify that s is a concave, decreasing, continuous function, and is the outer boundary function of set Z.

Step (iii). We decompose $[0, 1]^2$ into the following regions:

$$Z = \{z : 0 \le z_1 \le 0.63718 \text{ and } 0 \le z_2 \le s(z_1)\}; \quad \mathcal{A} = ([0, 0.16016] \times (0, 1]) \setminus Z$$
$$\mathcal{B} = ((0, 1] \times [0, 0.09] \setminus Z; \quad \mathcal{W} = [0, 1]^2 \setminus (Z \cup \mathcal{A} \cup \mathcal{B})$$

as illustrated in Figure 4. This is the canonical partition induced by set Z.

Step (iv). We claim that the canonical partition $Z \cup \mathcal{A} \cup \mathcal{B} \cup \mathcal{W}$ is well-formed with respect to μ. Condition 1 of Definition 15 is already satisfied given the definition of p^* in Step (ii). Moreover, recall that $S_{\text{top}}, S_{\text{right}} \subset \mathcal{Y}$ and, since $\mathcal{Y} \cup \{\vec{0}\}$ is a decreasing set, μ has negative density along these curves and all points below either curve, other than at the origin. Hence, $(\mu|_Z)_- \succeq_1 (\mu|_Z)_+$ and Condition 2 is also satisfied. Moreover, Conditions 4 and 5 are satisfied respectively by the definitions of S_{top} and S_{right} and the discussion surrounding these definitions in Step (i). Hence, the only non-trivial condition of Definition 15 that we need to verify is Condition 3, namely $\mu|_{\mathcal{W}} \succeq_2 0$. In fact, we can apply Lemma 5 to conclude the stronger dominance relation $\mu|_{\mathcal{W}} \succeq_1 0$. See Appendix E.2. Having verified all conditions of Definition 15 we can now apply Theorem 5 to conclude the following.

Example 2. *The optimal mechanism for selling two independent items whose values are distributed according to* $f_1(z_1) = \frac{z_1^2(1-z_1)^2}{B(3,3)}$ *and* $f_2(z_2) = \frac{z_2^2(1-z_2)^3}{B(3,4)}$ *has the following outcome for a bidder of type* (z_1, z_2) *in terms of the function* $s(\cdot)$ *defined above:*

- *If* $(z_1, z_2) \in Z$*, the bidder receives no goods and is charged 0.*
- *If* $(z_1, z_2) \in \mathcal{A}$*, the bidder receives item 1 with probability* $-s'(z_1)$*, item 2 with probability 1, and is charged* $s(z_1) - z_1 s'(z_1)$*.*
- *If* $(z_1, z_2) \in \mathcal{B}$*, the bidder receives item 1 with probability 1, item 2 with probability* $-1/s'(s^{-1}(z_2))$*, and is charged* $s^{-1}(z_2) - z_2/s'(s^{-1}(z_2))$*.*
- *If* $(z_1, z_2) \in \mathcal{W}$*, the bidder receives both items with probability 1 and is charged* $.71307$*.*

Recall that $s(z_1)$ is non-linear for $z_1 \in [0, .16016]$ and $z_1 \in [.62307, .63718]$. Hence, Example 2 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.

7.2.3 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 2, and Theorem 5 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 condition for grand bundling optimality (Theorem 3) for unbounded type distributions.

In Appendix G, we provide a short discussion of the modifications required to obtain an analog of Theorem 5 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 3, the optimal mechanism for selling two power-law items is a grand bundling mechanism. The canonical partition induced by the zero-set of the grand-bundling mechanism is degenerate (regions $\mathcal{A}$ and $\mathcal{B}$ are empty), and establishing the optimality of the mechanism amounts to establishing that the measure μ induced by the type distribution first-order stochastically dominates the 0 measure in region $\mathcal{W}$.

Example 3. *The optimal IC and IR mechanism for selling two items whose values are distributed independently according to the probability densities* $f_1(z_1) = 5/(1+z_1)^6$ *and* $f_2(z_2) = 6/(1+z_2)^7$ *respectively is a take-it-or-leave-it offer of the bundle of the two goods for price* $p^* \approx .35725$*.*

Example 4 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 zero-set of the mechanism is missing region $\mathcal{A}$, and possibly region $\mathcal{B}$ (if $\lambda_1 = \lambda_2$).

Example 4. *For all* $\lambda_1 \geq \lambda_2 > 0$*, the optimal IC and IR mechanism for selling two items whose values are distributed independently according to exponential distributions* f_1 *and* f_2 *with respective parameters* λ_1 *and* λ_2 *offers the following menu:*

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

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

$$\mu(\{y \in \mathbb{R}^2_{\geq 0} : y_1 + y_2 \leq p^* \text{ and } \lambda_1 y_1 + \lambda_2 y_2 \leq 2\}) = 0,$$

where μ is the transformed measure of the joint distribution.

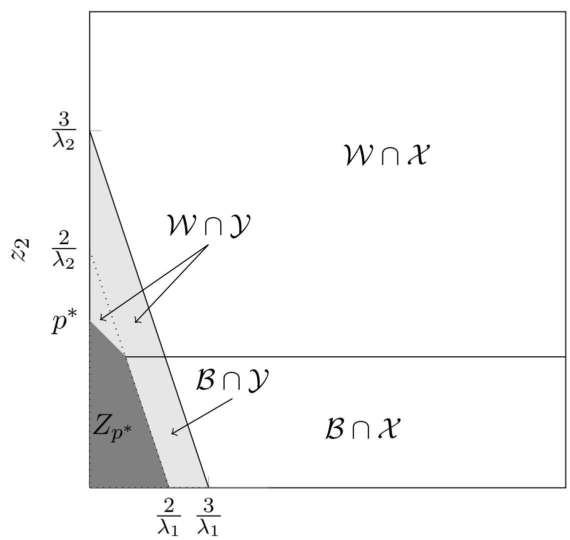


Figure 5: The canonical partition of $\mathbb{R}^n_{\geq 0}$ for the proof of Example 4. In this diagram, $p^* > 2/\lambda_1$. If $p^* \leq 2/\lambda_1$, $\mathcal{B}$ is empty. The positive part μ_+ of μ is supported inside $\mathcal{X} \cap \{\vec{0}\}$ while the negative part μ_- is supported within $Z_{p^*} \cup \mathcal{Y}$.

8 Proof of Strong Mechanism Design Duality

8.1 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 16 (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 Λ*, denoted Λ^*, is a map from the topological dual E^* of E to $\mathbb{R} \cup \{\infty\}$ given by*

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

Claim 6 (Fenchel-Rockafellar duality)**.** *Let E be a normed vector space, E^* its topological dual, and Θ, Ξ two convex functions on E taking values in $\mathbb{R} \cup \{+\infty\}$. Let Θ^*, Ξ^* be the Legendre-Fenchel transforms of Θ and Ξ respectively. Assume that there exists $z_0 \in E$ such that $\Theta(z_0) < +\infty$, $\Xi(z_0) < +\infty$ and Θ is continuous at z_0. Then*

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

Lemma 6. *Let X be a compact convex subset of $\mathbb{R}^n$, and let $\mu \in Radon(X)$ be such that $\mu(X) = 0$. Then*

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

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

Proof of Lemma 6: 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 $E^* = Radon(X \times X)$.

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

$$\Theta(f) = \begin{cases} 0 & \text{if } f(x,y) \geq -\|x-y\|_1 \text{ for all } x,y \in X \\ +\infty & \text{otherwise} \end{cases}$$
$$\Xi(f) = \begin{cases} \int_X \psi d\mu_- - \int_X \phi d\mu_+ & \text{if } f(x,y) = \psi(y) - \phi(x) \text{ for some } \psi,\phi \in \mathcal{U}(X) \\ +\infty & \text{otherwise.} \end{cases}$$

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

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

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

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

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

We claim therefore that

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

Indeed, if γ is a positive linear functional, then the result follows from monotonicity, since $\|x-y\|_1$ is the pointwise greatest function $\tilde{f}$ satisfying the constraint $\tilde{f}(x,y) \leq \|x-y\|_1$, and $\|x-y\|_1$ is continuous. Suppose instead that γ is a signed Radon measure which is not positive everywhere.

Then there exists a continuous nonnegative function $g : X \times X \to \mathbb{R}$ such that $\int g d\gamma = -\epsilon$ for some $\epsilon > 0$.[25] Since $g(x,y) \geq 0$, it follows that $-kg(x,y) \leq 0 \leq \|x-y\|_1$ for any $k \geq 0$. Therefore

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

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

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

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

We notice that $\Xi^*(\gamma) \geq 0$ for all $\gamma \in Radon(X \times X)$ by setting $\psi = \phi = 0$. In particular, we now compute for $\gamma \in Radon_+(X \times X)$:

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

We now apply Fenchel-Rockafellar duality:

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

□

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

8.2 From Two Convex Functions to One

Lemma 7. *Let $X = [0, M]^n$ for some $M \in \mathbb{R}_{\geq 0}$, and let $\mu \in Radon(X)$ such that $\mu(X) = 0$. Then*

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

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

Proof of Lemma 7: Given any feasible u for the right-hand side of Lemma 7, we observe that $\phi = \psi = u$ is feasible for the left-hand side, and therefore the left-hand side is at least as large as the right-hand side. It therefore suffices to prove the reverse direction of the inequality. Let ϕ and ψ be feasible for the left-hand side. Given ϕ, it is clear that ψ must satisfy $\psi(y) \geq \sup_x [\phi(x) - \|x-y\|_1]$.

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

We prove the following claim in Appendix F.1.

Claim 7. *$\bar{\psi} \in \mathcal{U}(X)$ and $\bar{\psi} \in \mathcal{L}_1(X)$.*

Since $\phi, \bar{\psi}$ are a feasible pair of functions for the left-hand side of Lemma 7, we know that ϕ satisfies the inequality $\phi(x) \leq \inf_y [\bar{\psi}(y) + \|x-y\|_1]$. We now set $\bar{\phi}(x) = \inf_y [\bar{\psi}(y) + \|x-y\|_1]$. It is clear that the value of the left-hand objective function under $\bar{\phi}, \bar{\psi}$ is at least as large as its value under $\phi, \bar{\psi}$.

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

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

where the last inequality holds since $\bar{\psi}(x) - \bar{\psi}(y) \leq \|x-y\|_1$. Therefore $\bar{\phi} = \bar{\psi}$, and thus $\bar{\phi} \in \mathcal{U}(X)$. Since $\bar{\phi}$ satisfies the inequality $\bar{\phi}(x) - \bar{\phi}(y) \leq \|x-y\|_1$ it is feasible for the right-hand side of Lemma 7, and the value of the right-hand objective under $\bar{\phi}$ is at least as large the value of the left-hand objective under ϕ, ψ. We notice finally that if ϕ, ψ are optimal for the left-hand side, then $\bar{\phi}$ is optimal for the right-hand side. □

8.3 Proof of Theorem 2

By combining Lemma 2, Lemma 6, and Lemma 7, we have

$$\begin{aligned} \inf_{\substack{\gamma \in Radon_+(X \times X) \\ \gamma_1 - \gamma_2 \succeq_1 \mu}} \int_{X \times X} \|x-y\|_1 d\gamma &\geq \sup_{u \in \mathcal{U}(X) \cap \mathcal{L}_1(X)} \int_X u d\mu \\ &= \sup_{\substack{\phi,\psi \in \mathcal{U}(X) \\ \phi(x)-\psi(y) \leq \|x-y\|_1}} \left(\int_X \phi d\mu_+ - \int_X \psi d\mu_- \right) = \inf_{\substack{\gamma \in Radon_+(X \times X) \\ \gamma_1 \succeq_{cvx} \mu_+ \\ \gamma_2 \preceq_{cvx} \mu_-}} \int_{X \times X} \|x-y\|_1 d\gamma(x,y). \end{aligned}$$

By Lemma 6, the last minimization problem above achieves its infimum for some γ^*. We notice that γ^* is also feasible for the first minimization problem above, and therefore the inequality is actually an equality and γ^* is optimal for the first minimization problem. In addition, since γ^* is feasible for the last minimization problem, it satisfies $\gamma_1^*(X) = \gamma_2^*(X) = \mu_+(X)$. 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 F.2.

References

[BB99] Yannis Bakos and Erik Brynjolfsson. Bundling information goods: Pricing, profits, and efficiency. *Management Science*, 45(12):1613–1630, 1999. 1

[DDT13] Constantinos Daskalakis, Alan Deckelbaum, and Christos Tzamos. Mechanism design via optimal transport. In *14th ACM Conference on Electronic Commerce (EC)*, 2013. 1, 2.1, 3, 4, 5, 16, 24, 7.2.3, G

[DFK11] Shahar Dobzinski, Hu Fu, and Robert D. Kleinberg. Optimal Auctions with Correlated Bidders are Easy. In *the 43rd ACM Symposium on Theory of Computing (STOC)*, 2011. A

[Dud02] R.M. Dudley. *Real Analysis and Probability*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2002. C.2, C.2

[FK12] Rafael M. Frongillo and Ian A. Kash. General truthfulness characterizations via convex analysis. *CoRR*, abs/1211.3043, 2012. A

[FKM11] Alessio Figalli, Young-Heon Kim, and Robert J McCann. When is multidimensional screening a convex program? *Journal of Economic Theory*, 146(2):454–478, 2011. 1

[GK14] Yiannis Giannakopoulos and Elias Koutsoupias. Duality and optimality of auctions for uniform distributions. In *the 15th ACM Conference on Electronic Commerce (EC)*, 2014. 1, 3

[HN13] Sergiu Hart and Noam Nisan. The menu-size complexity of auctions. In *14th ACM Conference on Electronic Commerce (EC)*, 2013. 1, 7.2.2

[MMW89] R. Preston McAfee, John McMillan, and Michael D. Whinston. Multiproduct monopoly, commodity bundling, and correlation of values. *The Quarterly Journal of Economics*, pages 371–383, 1989. 1

[MV06] A.M. Manelli and D.R. Vincent. Bundling as an optimal selling mechanism for a multiple-good monopolist. *Journal of Economic Theory*, 127(1):1–35, 2006. 1, 2.1, 2.1, 2.1, 3, 5, 7.2, 7.2.1

[MV07] A. M. Manelli and D. R. Vincent. Multidimensional Mechanism Design: Revenue Maximization and the Multiple-Good Monopoly. *Journal of Economic Theory*, 137(1):153–185, 2007. 1

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

[Pav11] Gregory Pavlov. Optimal mechanism for selling two goods. *The B.E. Journal of Theoretical Economics*, 11(1):1–35, February 2011. (document), 1, 4, 4, 5

[RC98] Jean-Charles Rochet and Philippe Choné. Ironing, sweeping, and multidimensional screening. *Econometrica*, 66(4):783–826, July 1998. 1, 2.1, 2.2.2, 3

[Roc87] Jean-Charles Rochet. A necessary and sufficient condition for rationalizability in a quasi-linear context. *Journal of Mathematical Economics*, 16(2):191–200, April 1987. 1, 2.1, 2.1

[Rod87] José-Francisco Rodrigues. *Obstacle Problems in Mathematical Physics*. North-Holland, 1987. A

[RS03] Jean-Charles Rochet and Lars A Stole. The economics of multidimensional screening. *Econometric Society Monographs*, 35:150–197, 2003. 1

[SS10] M. Shaked and J.G. Shanthikumar. *Stochastic Orders*. Springer Series in Statistics. Springer, 2010. 6, C.1, C.2

[Tha04] J. Thanassoulis. Haggling over substitutes. *Journal of Economic Theory*, 117(2):217–245, 2004. 1

[Vil09] C. Villani. *Optimal Transport: Old and New*. Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen mit besonderer Berücksichtigung der Anwendungsgebiete. Springer Berlin Heidelberg, 2009. 1, 3.2, 8.1

A Omitted Details from Section 2

While our definitions of individual rationality and incentive compatibility apply to mechanisms over an arbitrary type space $Z \subseteq \mathbb{R}^n_{\geq 0}$, it is well-known (see, e.g., similar results in [DFK11, FK12]) that we can extend a mechanism's type space to any superset of Z without violating the IC or IR properties. Our proof constructs a new mechanism $\mathcal{M}'$ providing the new types a choice from the (closure of the) of price-allocation pairs offered by $\mathcal{M}$.

Fact 1 (Domain Extension)**.** *Let $\mathcal{M} = (\mathcal{P}, \mathcal{T})$ be an IC and IR mechanism over type space $Z \subseteq \mathbb{R}^n_{\geq 0}$, and let $X \supseteq Z$ be any subset of $\mathbb{R}^n_{\geq 0}$ containing Z. Then there exists an IC and IR mechanism $\mathcal{M}' = (\mathcal{P}', \mathcal{T}')$ over type space X such that $\mathcal{P}(z) = \mathcal{P}'(z)$ and $\mathcal{T}(z) = \mathcal{T}'(z)$ for all $z \in Z$.*

Proof of Fact 1: Let $\mathcal{O} = \{(\mathcal{P}(z), \mathcal{T}(z)) : z \in Z\} \cup \{(0^n, 0)\}$ be the set of all possible outcome-price pairs of the mechanism $\mathcal{M}$, along with the "opt-out" option which allocates no items and charges nothing. We now define the set $\tilde{\mathcal{O}}$ to be the closure of $\mathcal{O}$, under the standard topology on $\mathbb{R}^{n+1}$.

We define the mechanism $\mathcal{M}'$ as follows:

$$(\mathcal{P}'(z'), \mathcal{T}'(z')) = \begin{cases} (\mathcal{P}(z'), \mathcal{T}(z')), & \text{if } z' \in Z \\ \arg\max_{(p,t) \in \tilde{\mathcal{O}}} (z' \cdot p - t), & \text{otherwise.} \end{cases}$$

where the above argmax is chosen arbitrary if it is not unique. Mechanism $\mathcal{M}'$ behaves as $\mathcal{M}$ on all declared types $z' \in Z$, and allows declared types outside of Z to choose any outcome in $\tilde{\mathcal{O}}$ which would maximize the bidder's utility. We argue first that the above argmax is never empty: Notice that the set $\{\mathcal{T}(z) : z \in Z\}$ is bounded from below by some value $-b$, as otherwise $\mathcal{M}$ would make arbitrarily high payouts to the players in Z, meaning that no player would have a best strategy in $\mathcal{M}$ (since, regardless of his type, he could always play a different strategy in which the mechanism pays him significantly money) violating the IC property. Furthermore, for any $z' \in X \setminus Z$ we need only consider allocations with $t \leq \|z'\|_1$. Thus we are optimizing a continuous function over a compact subset of (p, t) values, namely $\mathcal{O} \cap ([0, 1] \times [-b, \|z'\|_1])$. The argmax is therefore nonempty. It is straightforward to show that $\mathcal{M}'$ is IC and IR, so we omit these verifications. □

We now note in Remark 6 that a non-decreasing function being 1-Lipschitz with respect to the ℓ_1 norm implies that its gradient exists almost everywhere. This property is useful for Claim 1. The proof of this remark is an application of Rademacher's Theorem to the interior of the set X.

Remark 6. Let X be a nonempty subset of $\mathbb{R}^n$ whose boundary has Lebesgue measure 0. Then every non-decreasing function $u : X \to \mathbb{R}$ which is 1-Lipschitz with respect to the ℓ_1 norm satisfies $\nabla u \in [0, 1]^n$ almost everywhere.

Proof of Remark 6: Suppose that u is non-decreasing and 1-Lipschitz. Since the boundary of X has measure 0, it suffices to show that $\nabla u \in [0, 1]^n$ almost everywhere on $int(X)$, the interior of X. We now apply Rademacher's Theorem, which states that a Lipschitz continuous function mapping an open subset of $\mathbb{R}^n$ into $\mathbb{R}$ is differentiable almost everywhere, to conclude that ∇u exists at almost all $x \in int(X)$. For any point x where ∇u exists, and any coordinate vector e_i, we have $\frac{\partial u}{\partial z_i}(x) = \lim_{\epsilon \to 0} \frac{u(x+\epsilon e_i) - u(x)}{\epsilon}$. We now apply the 1-Lipschitz condition and the fact that u is non-decreasing to bound $0 \leq u(x + \epsilon e_i) - u(x) \leq \epsilon$ to conclude that $\nabla u \in [0, 1]^n$ almost everywhere. □

Proof of Claim 2: Let u be the utility function of an IC and IR mechanism $\mathcal{M} = (\mathcal{P}, \mathcal{T})$ over type space U, and let $\mathcal{P}', \mathcal{T}'$ be an extension of $\mathcal{M}$ to X, as constructed in Fact 1. The expected revenue of $\mathcal{M}$ is given by $\int_X f(z)\mathcal{T}'(z)dz$. We now apply Claim 1 to write the expected revenue as

$$\begin{aligned}\int_X \mathcal{T}'(z)f(z)dz &= \int_U \mathcal{T}(z)f(z)dz = \int_U \left[\nabla u(z)\cdot z - u(z)\right]f(z)dz \\ &= \sum_i \left[\int_U \left(\frac{\partial u}{\partial z_i}z_i f + u\frac{\partial z_i f}{\partial z_i} - u\frac{\partial z_i f}{\partial z_i}\right)dz\right] - \int_U ufdz \\ &= \sum_i \left[\int_U \left(\frac{\partial u}{\partial z_i}z_i f + u\frac{\partial z_i f}{\partial z_i} - uz_i\frac{\partial f}{\partial z_i}\right)dz\right] - (n+1)\int_U ufdz \\ &= \sum_i \left[\int_U \left(\frac{\partial u}{\partial z_i}z_i f + u\frac{\partial z_i f}{\partial z_i}\right)dz\right] - \int_U (u\nabla f\cdot z + (n+1)uf)dz\end{aligned}$$

We now apply a high-dimensional analog of integration by parts (see Theorem 6.1 of Chapter 3 of [Rod87]) to write $\int_U \left(\frac{\partial u}{\partial z_i}z_i f + u\frac{\partial z_i f}{\partial z_i}\right)dz$ as a surface integral $\int_{\partial U} ufz_i\hat{n}_i ds$. This theorem applies because $u(z)$ and $z_i f(z)$ are both Lipschitz functions and U is a bounded Lipschitz domain. The result follows by noting that $\sum_i \left[\int_{\partial U} ufz_i\hat{n}_i ds\right] = \int_{\partial U}(z\cdot\hat{n})ufds$. □

Proof of Claim 3: Clearly, any feasible $u : X \to \mathbb{R}$ for the left-hand side is feasible for the right-hand side with $\int_X ud\nu \le \int_X ud\mu$ (since $u(z_0) \ge 0$). Thus the supremum value of the left-hand side is at most as large as the supremum value of the right-hand side.

For the reverse direction, let u be feasible for the right-hand side, and define the function $\bar{u} : X \to \mathbb{R}$ by $\bar{u}(x) \triangleq \max\{u(x) - u(z_0), 0\}$. It is straightforward to show that $\bar{u}$ is feasible for the left-hand. Furthermore, since u is non-decreasing and continuous, by our choice of z_0 we have $\bar{u}(z) = u(z) - u(z_0)$ for all $z \in U \cup \partial U \cup \{z_0\}$. Observing that μ and ν are both supported entirely within $U \cup \partial U \cup \{z_0\}$ and $U \cup \partial U$, respectively, we now compute

$$\begin{aligned}\int_X \bar{u}d\nu = -\bar{u}(z_0) + \int_X \bar{u}d\mu = 0 + \int_X (u(z) - u(z_0))d\mu \\ = \int_X ud\mu - u(z_0)\mu(X) = \int_X ud\mu\end{aligned}$$

and thus the supremum value of the left-hand side is at least as large as the supremum value of the right-hand side. □

B Omitted Details from Example 1

We present omitted details from the proof of Example 1, namely in the construction of γ^Y

Since $\mu_+|_Y + \alpha_+$ only assigns mass to the upper boundary of Y, to show that γ^Y can be constructed so that all mass is transported "vertically downwards" we need only verify that $\mu_+|_Y + \alpha_+$ and $\mu_-|_Y + \alpha_-$ assign the same density to any vertical "strip" in Y. Indeed,

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

and, for all $z_1 \pm \epsilon \in (4, 8]$, we compute the following, using the fact that the surface area of

$Y \cap ([z_1-\epsilon, z_1+\epsilon] \times [4,7])$ is $2\epsilon \cdot \left(\frac{z_1}{2}-1\right)$:

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

C Proof of Grand Bundling Optimality Theorem

C.1 Lemmas on Stochastic Dominance

Before we can prove Theorem 3, we need to show several lemmas. The most important of these is Lemma 12 in Section C.2, which we expect can be a key tool for adapting the characterization and proof of Theorem 3 to other classes of mechanisms. Before stating this lemma, we first show two simple results.

Lemma 8. *Let A and B be vector random variables with values in $[0, M]^n$, and suppose that $\mathbb{E}[\|A\|_1] = \mathbb{E}[\|B\|_1]$. Then $A \succeq_{cvx} B$ if and only if $B \succeq_2 A$.*

Proof of Lemma 8: Suppose that $A \succeq_{cvx} B$. Let $h : \mathbb{R}^n \to \mathbb{R}$ be an arbitrary non-decreasing concave function, and let

$$m = \sup_{x \neq y \in [0,M]^n} \frac{h(x) - h(y)}{\|x - y\|_1}.$$

Since h is defined on an open subset containing $[0, M]^n$, we know that h is Lipschitz continuous on $[0, M]^n$, and therefore m is finite. Thus, the function

$$g(x) = -h(x) + m\|x\|_1$$

is convex, continuous, and non-decreasing. Therefore

$$\mathbb{E}[h(A) - h(B)] = \mathbb{E}[m\|A\|_1 - m\|B\|_1] - \mathbb{E}[g(A) - g(B)] \leq 0$$

and thus $A \preceq_2 B$. The proof of the other direction of the lemma is analogous. □

Lemma 9. *Let X be a convex bounded subset of $\mathbb{R}^n$ and let $\alpha, \beta \in Radon_+(X)$ such that $\alpha \preceq_2 \beta$. Then there exists $\theta \in Radon_+(X)$ such that $\alpha \preceq_1 \theta$, $\theta \preceq_2 \beta$, and $\int \|x\|_1 d\theta = \int \|x\|_1 d\beta$.*

Proof of Lemma 9: This proof follows from Theorem 7.A.3 and Theorem 4.A.6 of [SS10] applied to second order dominance, which state that for vector-valued random variables $A \preceq_2 B$ there exists a vector valued random variable T such that $A \preceq_1 T \preceq_{cv} B$, where $T \preceq_{cv} B$ means that $\mathbb{E}[g(T)] \leq \mathbb{E}[g(B)]$ for all concave (but not necessarily monotonic) functions $g : \mathbb{R}^n \to \mathbb{R}$.

In particular, $T \preceq_{cv} B$ implies both that $T \preceq_2 B$ and that $\mathbb{E}[\|T\|_1] = \mathbb{E}[\|B\|_1]$. We notice that, since the proof of this lemma defines T by taking conditional expectations of A and B, the values of T always lie within the convex set X. The result follows by taking an analogous statement for positive Radon measures. □

C.2 Probabilistic Lemmas

The goal of this section is to prove Lemma 12, which will be very useful in our proof of Theorem 3. 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 10. *Let X and Y be random vectors. Then $X \preceq_{cvx} Y$ if and only if there exist random vectors $\hat{X}$ and $\hat{Y}$, defined on the same probability space, such that $\hat{X} =_{st} X$, $\hat{Y} =_{st} Y$, and $\mathbb{E}[\hat{Y}|\hat{X}] \geq \hat{X}$ almost surely, where the final inequality is componentwise and where $=_{st}$ denotes equality in distribution.*

Similarly, $X \preceq_2 Y$ if and only if there exist random vectors $\hat{X}$ and $\hat{Y}$, defined on the same probability space, such that $\hat{X} =_{st} X$, $\hat{Y} =_{st} Y$, and $\mathbb{E}[\hat{X}|\hat{Y}] \leq \hat{Y}$ almost surely.

We now state a standard 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.

Claim 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 $\mathbb{E}[u(V)] \geq u(\mathbb{E}[V])$. Furthermore, equality holds if and only if, for every a in the subdifferential of u at $\mathbb{E}[V]$, the equality $u(V) = a \cdot (V - \mathbb{E}[V]) + u(\mathbb{E}[V])$ holds almost surely.*

The following lemma is a conditional variant of Claim 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 12, the main result of this subsection.

Lemma 11. *Let $(\Omega, \mathcal{A}, P)$ be a probability space, V be a random variable on Ω with values in $[0, M]^n$, and $u : [0, M]^n \to \mathbb{R}$ be convex and Lipschitz continuous. Let $\mathcal{C}$ be any sub-σ-algebra of $\mathcal{A}$ and suppose that $\mathbb{E}[u(V)|\mathcal{C}] = u(\mathbb{E}[V|\mathcal{C}])$ almost-surely. Then for almost all $x \in \Omega$ the equality $u(y) = a_{y_x} \cdot (y - y_x) + u(y_x)$ holds almost surely with respect to the law $P_{V|\mathcal{C}}(\cdot, x)$, 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 11: 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 $[0, M]^n$. The multivariate conditional Jensen's inequality states that, almost surely, $\mathbb{E}[V|\mathcal{C}] \in C$ and $\mathbb{E}[u(V)|\mathcal{C}] \geq u(\mathbb{E}[V|\mathcal{C}])$. The proof of Theorem 10.2.7 in [Dud02] furthermore shows that the following two equalities hold:

$$\mathbb{E}[V|\mathcal{C}](x) = \int_{[0,M]^n} y P_{V|\mathcal{C}}(dy, x); \qquad \mathbb{E}[u(V)|\mathcal{C}](x) = \int_{[0,M]^n} u(y) P_{V|\mathcal{C}}(dy, x).$$

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

We now present Lemma 12. Very roughly, this lemma states that for random variables X and Y with $X \preceq_{cvx} Y$ if it holds that $u(X) = u(Y)$ for some convex function u, then there exists a coupling between X and Y with several desirable properties, including that points are only matched if u shares a subgradient at these points.

Lemma 12. *Let X and Y be vector random variables with values in $[0,M]^n$ such that $X \preceq_{cvx} Y$, and let $u : [0,M]^n \to \mathbb{R}$ be 1-Lipschitz with respect to the ℓ_1 norm, convex, and monotonically non-decreasing. Suppose that $\mathbb{E}[u(X)] = \mathbb{E}[u(Y)]$ and that $g : [0,M]^n \to [0,1]^n$ is a measurable function such that for all $z \in [0,M]^n$, $g(z)$ is a subgradient of u at z.*

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

- $u(\hat{Y}) = u(\hat{X}) + g(\hat{X}) \cdot (\hat{Y} - \hat{X})$
- *$g(\hat{X})$ is a subgradient of u at $\hat{Y}$.*
- *$\mathbb{E}[\hat{Y}|\hat{X}]$ is componentwise greater than $\hat{X}$*
- $u(\mathbb{E}[\hat{Y}|\hat{X}]) = u(\hat{X})$.

Proof of Lemma 12: By Lemma 10, there exist random variables $\hat{X} =_{st} X$ and $\hat{Y} =_{st} Y$ such that $\mathbb{E}[\hat{Y}|\hat{X}]$ is componentwise greater than or equal to $\hat{X}$ almost surely. We have

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

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

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

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

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

and therefore $a_x \cdot \mathbb{E}[\hat{Y}|_{\hat{X}=x}] = a_x \cdot x$, by non-negativity of the subgradient. Furthermore, for any point $z \in [0,M]^n$,

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

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

Since $\mathbb{E}[\mathbb{E}[u(\hat{Y})|\hat{X}]] = \mathbb{E}[u(\mathbb{E}[\hat{Y}|\hat{X}])]$, by Jensen's inequality it follows that $\mathbb{E}[u(\hat{Y})|\hat{X}] = u(\mathbb{E}[\hat{Y}|\hat{X}])$ almost surely. By Lemma 11, it therefore holds for almost all values x of $\hat{X}$ that the equality

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

holds $\hat{Y}|_{\hat{X}=x}$ almost surely.

Lastly, we will show that, almost surely, a_x is a subgradient of u at $\hat{Y}|_{\hat{X}=x}$. Indeed, for any $p \in [0,M]^n$, and almost all values of x we have

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

□

C.3 Proof of Grand Bundling Theorem

Our goal in this section is to prove Lemma 3, which immediately implies Theorem 3. We begin by proving the following corollary of Lemma 12. Roughly speaking, this corollary gives a useful property of Radon measures: if one measure convexly dominates another and if the grand bundling utility function u_p has the same expectation under both of these measures, then appropriate stochastic dominance relations hold when restricted to the regions Z and W.

Corollary 2. *Let $X = [0, M]^n$, and let A, B be random variables with values in $[0, M]^n$ such that $A \preceq_{cvx} B$.*

For $p \in \mathbb{R}_{\geq 0}$, define the function $u_p : X \to R_{\geq 0}$ by $u_p(x) = \max\{\|x\|_1 - p, 0\}$ and define the regions $Z, P, W \subset X$ by

$$Z = \{x \in X : \|x\|_1 \leq p\}\,; \quad P = \{x \in X : \|x\|_1 = p\}\,; \quad W = \{x \in X : \|x\|_1 \geq p\}\,.$$

If $\mathbb{E}[u_p(A)] = \mathbb{E}[u_p(B)]$, then there exist $\hat{A} =_{st} A$ and $\hat{B} =_{st} B$ such that $\hat{A} \leq \mathbb{E}[\hat{B}|\hat{A}]$ componentwise holds almost surely and

$$\Big((\hat{A} \in P) \cap (\hat{B} \in P)\Big) \cup \Big((\hat{A} \in Z) \cap (\hat{B} \in Z)\Big) \cup \Big((\hat{A} \in W) \cap (\hat{B} \in W)\Big)$$

holds almost surely. Furthermore,

$$A \cdot \mathbb{I}_{A \in Z} \preceq_{cvx} B \cdot \mathbb{I}_{B \in Z}, \quad A \cdot \mathbb{I}_{A \in P} \preceq_{cvx} B \cdot \mathbb{I}_{B \in P}, \quad \textit{and} \quad A \cdot \mathbb{I}_{A \in W} \preceq_{cvx} B \cdot \mathbb{I}_{B \in W}.$$

Proof of Corollary 2: Select $\hat{A}$ and $\hat{B}$ as in Lemma 12, taking $u = u_p$ and $g(x)$ to be 0^n if $x \in Z$, $(\frac{1}{2})^n$ if $z \in P$, and 1^n if $z \in W$. The result that $\hat{B}$ lies in the same region as $\hat{A}$ follows from the property that $g(\hat{A})$ is a subgradient of u_p at $\hat{B}$ almost surely.

The convex dominance conditions follow from Strassen's theorem for convex dominance, by observing that, for any region $R = Z$, P, or W, the coupling between $\hat{A}$ and $\hat{B}$ satisfies

$$\mathbb{E}[\hat{B} \cdot \mathbb{I}_{\hat{A},\hat{B} \in R} | \hat{A} \cdot \mathbb{I}_{\hat{A},\hat{B} \in R}] \geq \hat{A} \cdot \mathbb{I}_{\hat{A},\hat{B} \in R}$$

almost surely. Finally, it is obvious that a relation such as $A \cdot \mathbb{I}_{A \in Z} \preceq_{cvx} B \cdot \mathbb{I}_{B \in Z}$ holds if and only if $\hat{A} \cdot \mathbb{I}_{\hat{A} \in Z} \preceq_{cvx} \hat{B} \cdot \mathbb{I}_{\hat{B} \in Z}$, as $\hat{A} =_{st} A$ and $\hat{B} =_{st} B$. □

Remark 7. The above corollary can be extended to many functions u other than the one described above. Informally, Lemma 12 allows us to couple random variables $A \preceq_{cvx} B$ into random variables $\hat{A}$ and $\hat{B}$ such that, almost surely, the gradients of u at $\hat{A}$ and $\hat{B}$ coincide. Such a result aids in the decomposition of a single stochastic dominance condition into several stochastic dominance conditions over regions in which the gradients of u are constant.

We will now prove Lemma 3 (and hence Theorem 3) by showing that the first and second conditions are equivalent.

Suppose that μ satisfies the second condition of the theorem. We will construct an appropriate measure $\gamma \in Radon_+(W \times W)$. Since $\mu_-|_W \preceq_2 \mu_+|_W$, by Lemma 9 there exists a measure $\theta \in Radon_+(W)$ such that

$$\mu_-|_W \preceq_1 \theta \preceq_2 \mu_+|_W \quad \text{and} \quad \int \|x\|_1 d\theta = \int \|x\|_1 d\mu_+|_W.$$

Since $\theta \succeq_1 \mu_-|_W$, by Strassen's theorem there exists $\gamma \in \Gamma(\theta, \mu_-|_W)$ such that $x \geq y$ holds $\gamma(x, y)$ almost surely.

We now verify the following points

- We have $\gamma_1 - \gamma_2 = \theta - \mu_-|_W \preceq_2 \mu|_W$. Furthermore, by Lemma 8, we know that $\gamma_1 - \gamma_2 \succeq_{cvx} \mu|_W$. Since $0 \succeq_{cvx} \mu|_Z = \mu|_{Z\setminus P}$, we have $\gamma_1 - \gamma_2 \succeq_{cvx} \mu|_{Z\setminus P} + \mu_W = \mu$.

- We first note that

$$\int \|x\|_1 d(\gamma_1 - \gamma_2) = \int \|x\|_1 d(\theta - \mu_-|_W) = \int \|x\|_1 d\mu|_W.$$

  Furthermore, since $(\gamma_1 - \gamma_2)(W) = \mu(W) = 0$, we have

$$\int (\|x\|_1 - p) d(\gamma_1 - \gamma_2) = \int (\|x\|_1 - p) d\mu|_W.$$

  Since $u_p(x) = 0$ off of W, $u_p(x) = \|x\|_1 - p$ on W, and $\gamma \in \mathit{Radon}_+(W \times W)$,

$$\int (\|x\|_1 - p) d(\gamma_1 - \gamma_2) = \int u_p d(\gamma_1 - \gamma_2) = \int u_p d\mu|_W = \int u_p d\mu$$

- Finally, by construction, it holds $\gamma(x, y)$ almost surely that $x \geq y$, and therefore $u_p(x) - u_p(y) = \|x\|_1 - \|y\|_1 = \|x - y\|_1$ holds $\gamma(x, y)$ almost surely.

Now suppose that $\gamma \in \mathit{Radon}_+(X \times X)$ satisfies condition 1. Since $\gamma_1 + \mu_- \succeq_{cvx} \gamma_2 + \mu_+$, by Corollary 2 we know that

$$(\gamma_1 + \mu_-)\,|_Z \succeq_{cvx} (\gamma_2 + \mu_+)\,|_Z \quad \text{and} \quad (\gamma_1 + \mu_-)\,|_W \succeq_{cvx} (\gamma_2 + \mu_+)\,|_W.$$

Since $u_p(x) - u_p(y) = \|x - y\|_1$ holds $\gamma(x, y)$-almost surely, we know that almost surely either `(i)` $x, y \in Z$ and $x = y$ or `(ii)` $x, y \in W$ with $x \geq y$ coordinatewise.

We can therefore (uniquely) decompose $\gamma = \zeta + \eta$, where $\zeta \in \mathit{Radon}_+(Z \setminus P, Z \setminus P)$ and $\eta \in \mathit{Radon}_+(W, W)$ such that $x = y$ holds $\zeta(x, y)$ almost surely and $x \geq y$ coordinatewise holds $\eta(x, y)$ almost surely. We now make several claims

- We will show first that $\mu(W) = 0$. Since $\gamma_1|_W = \eta_1$ and $\gamma_2|_W = \eta_2$:

$$\begin{aligned} \gamma_1|_W + \mu_-|_W &\succeq_{cvx} \gamma_2|_W + \mu_+|_W \\ \eta_1 + \mu_-|_W &\succeq_{cvx} \eta_2 + \mu_+|_W. \end{aligned}$$

  Since convex dominance only holds between measures with equal total mass, $\eta_1(W) - \eta_2(W) = \mu(W)$. But since $\eta \in \mathit{Radon}_+(W, W)$, we know that $\eta_1(W) = \eta_2(W)$, and therefore $\mu(W) = 0$.

- We claim next that $\eta_1|_P = \eta_2|_P$. In particular, since $\gamma_1|_Z = \zeta_1 + \eta_1|_P$ and $\gamma_2|_Z = \zeta_2 + \eta_2|_P$, we have

$$\begin{aligned} \gamma_1|_Z + \mu_-|_Z &\succeq_{cvx} \gamma_2|_Z + \mu_+|_Z \\ \zeta_1 + \eta_1|_P + \mu_-|_Z &\succeq_{cvx} \zeta_2 + \eta_2|_P + \mu_+|_Z. \end{aligned}$$

  Since $x = y$ holds $\zeta(x, y)$ almost surely, it follows that $\zeta_1 = \zeta_2$, and thus $\eta_1|_P + \mu_-|_Z \succeq_{cvx} \eta_2|_P + \mu_+|_Z$. Convex dominance only holds between two measures with equal total mass, and therefore

$$\eta_1|_P(Z) - \eta_2|_P(Z) = \mu(Z) = \mu(X) + \mu(P) - \mu(W) = 0.$$

  Since $x \geq y$ coordinatewise holds $\eta(x, y)$ almost surely, and since P does not contain any $x \neq y$ such that $x \geq y$ coordinatewise, if $\eta_1|_P \neq \eta_2|_P$ we would have $\eta_1|_P(Z) < \eta_2|_P(Z)$, which is a contradiction. Therefore $\eta_1|_P = \eta_2|_P$

- We now claim that $\mu_-|_Z \succeq_{cvx} \mu_+|_Z$. We compute, using the fact that $\gamma_1|_Z = \zeta_1 + \eta_1|_P$ and $\gamma_2|_Z = \zeta_2 + \eta_2|_P$,

$$\gamma_1|_Z + \mu_-|_Z \succeq_{cvx} \gamma_2|_Z + \mu_+|_Z$$
$$\zeta_1 + \eta_1|_P + \mu_-|_Z \succeq_{cvx} \zeta_2 + \eta_2|_P + \mu_+|_Z$$

  Since $\eta_1|_P = \eta_2|_P$ and $\zeta_1 = \zeta_2$, we have $\mu_-|_Z \succeq_{cvx} \mu_+|_Z$.

- We claim finally that $\mu_-|_W \preceq_2 \mu_+|_W$. Since $u = 0$ on Z, we know

$$\int_W (\|x\|_1 - p)d\gamma_1 - \int_W (\|x\|_1 - p)d\gamma_2 = \int_W (\|x\|_1 - p)d\mu.$$

  Since $\gamma_1|_W = \eta_1$, $\gamma_2|_W = \eta_2$, $\eta_1(W) = \eta_2(W)$, and $\mu(W) = 0$, we have

$$\int_W \|x\|_1 d(\eta_1 - \eta_2) = \int_W \|x\|_1 d\mu.$$

  In addition, we have $\eta_1 - \eta_2 \succeq_{cvx} \mu|_W$. This implies by Lemma 8 that $\mu|_W \succeq_2 \eta_1 - \eta_2$. Furthermore, since $\eta_1 - \eta_2 \succeq_1 0$ (by Strassen's theorem, as $x \geq y$ holds $\eta(x, y)$ almost surely) we have $\mu|_W \succeq_2 0$, and thus $\mu_+|_W \succeq_2 \mu_-|_W$.

C.4 Example of Grand Bundling Optimality in the Hypercube

In this appendix we complete the proof of Theorem 4.
Proof of Lemma 4: We define the mapping $\varphi : A \to B$ by $\varphi(x) = y$, where

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

We first claim that φ is a bijection. As x_n ranges from 0 to $1 - \left(\frac{\rho-1}{\rho}\right)^{1/(n-1)}$, we see that y_1 ranges from 1 to 0, and thus there is a bijection between valid y_1 values and valid x_n values. Furthermore, for any fixed y_1 and x_n, there is a bijection between x_i and y_i for $i = 2, \ldots, n-1$. (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, φ is a bijection between A and B.

Next, we claim that for any $x \in A$, it holds that x is componentwise at least as large as $\varphi(x)$. Since $x_1 = 1$, it trivially holds that $x_1 \geq (\varphi(x))_1$. Fix a value of x_n (and hence of y_1), and consider the bijection $g : [x_n, 1] \to [0, y_1]$ given by $g(z) = y_1(z - x_n)/(1 - x_n)$. We must show that $z - g(z) \geq 0$ for all $z \in [x_n, 1]$. This follows from noticing that $z - g(z)$ is a linear function of z and both $x_n - g(x_n) = x_n$ and $1 - g(1) = 1 - y_1$ are nonnegative.

We now show that φ scales surface measure of every measurable $S \subset A$ by a factor of $1/\rho$. Instead of directly analyzing surface measures, it suffices to prove that the function $\varphi' : W \to W$ scales volumes by ρ, 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 φ'. We notice that the only non-zero entries of J are the diagonals and the rightmost column. In particular, J is upper triangular, and therefore its determinant is the product of its diagonal entries. We therefore compute

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

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 $x_n \geq 1-\left(\frac{\epsilon^{n-1}+\rho-1}{\rho}\right)^{1/(n-1)}$. □

Proof of Theorem 4: We now complete the proof of Theorem 4. 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 $(c, c, \dots, c)$.
- Mass of $-(n+1)$ uniformly distributed throughout the interior.
- Mass of -c distributed on each surface $x_i = c$ of the hypercube.
- Mass of $c+1$ distributed on each surface $x_i = c+1$ of the hypercube.

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

For any $h \in (0,1)$, define the region $Z(h) = \{x \in [0,1]^n : \|x\|_1 \leq h\}$. For any fixed c_0, it holds that $\mu_+^{c_0}(Z(h)) = 1$ for all $h \in (0,1)$ and there exists a small enough $h' > 0$ such that $\mu_-^{c_0}(Z(h')) < 1$. Since for this fixed h' it holds that $\mu_-^c(Z(h'))$ increases with c (and becomes arbitrarily large as c becomes large), there must exist a $c' > c_0$ such that $\mu_-^{c'}(Z(h')) = 1$, and thus $\mu^{c'}(Z(h')) = 0$. We can therefore pick a decreasing function $p^* : \mathbb{R}_{\geq 0} \to (0,1)$ such that, for all sufficiently large c, $\mu^c(Z(p^*(c))) = 0$.[26] It follows that $p^*(c) \to 0$ as $c \to \infty$.

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

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

We notice that $\mu_+^c(Z_c \cap W_c) = \mu_-^c(Z_c \cap W_c) = 0$. By construction, for large enough c we have $\mu^c(Z_c) = 0$. In addition, the only positive mass in Z_c is at the origin, and thus $\mu_-^c|_{Z_c} \succeq_1 \mu_+^c|_{Z_c}$.

To apply Theorem 3, it remains to show that, for sufficiently large c, $\mu_+^c|_{W_c} \succeq_2 \mu_-^c|_{W_c}$. To prove this, we will partition W_c into $2(n!+1)$ disjoint[27] regions, $P_0, P_{\sigma_1}, \dots, P_{\sigma_{n!}}$ and $N_0, N_{\sigma_1}, \dots, N_{\sigma_{n!}}$, where σ_j is a permutation of $1, \dots, n$. This partition will be such that $\cup_j P_j$ contains the entire support of $\mu_+^c|_{W_c}$ and $\cup_j N_j$ contains the entire support of $\mu_-^c|_{W_c}$. We will show that $\mu_+^c|_{P_j} \succeq_2 \mu_-^c|_{N_j}$ for all j, thereby proving $\mu_+^c|_{W_c} \succeq_2 \mu_-^c|_{W_c}$.

[26] 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.

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

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

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

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

Denote $N_\sigma \triangleq N'_\sigma \setminus Z_c$ and $P_\sigma \triangleq \varphi^{-1}(N_\sigma)$. By construction, φ is a bijection between P_σ and N_σ, preserving the respective the measures μ_+^c and μ_-^c, such that for all $x \in P_\sigma$, x is componentwise at least as large as $\varphi(x)$. Therefore, by Strassen's theorem, $\mu_+^c|_{P_\sigma} \succeq_1 \mu_-^c|_{N_\sigma}$. Lastly, we define

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

P_0 consists of all points on the outer surface of the hypercube which have not yet been matched to any N_σ, and N_0 consists of all points on which μ_-^c is nontrivial which have not yet been matched.[28] It therefore remains only to show that $\mu_+^c|_{P_0} \succeq_2 \mu_-^c|_{N_0}$.

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

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

By construction, $\mu_-^c|_{N_0}$ and $\mu_+^c|_{P_0}$ have the same total mass. Consider independent random variables X and Y corresponding to $\mu_-^c|_{N_0}$ and $\mu_+^c|_{P_0}$, respectively, where we scale both measures so that they are probability distributions. By Lemma 10, it suffices to show that for sufficiently large c, $Y \geq \mathbb{E}[X]$ almost surely.[29] Since $\mu_+^c|_{P_0}$ is supported on P_0, we need only show that all coordinates of $\mathbb{E}[X]$ are less than $3/4$. We recall that μ_-^c assigns a total mass of $n+1$, distributed uniformly, to the interior of the hypercube. As c gets large, $p^*(c)$ approaches 0, and thus $\mu_-^c(Z_c \cap (0,1)^n)/\mu_-^c((0,1)^n) \to 0$. For large c, therefore, $\mathbb{E}[X]$ becomes arbitrarily close to the center of the hypercube, which is the point with all coordinates equal to $1/2$. Therefore we have $\mu_+^c|_{P_0} \succeq_2 \mu_-^c|_{N_0}$. □

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

[29] 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.

D Supplementary Material for Section 6

Proof of Claim 4: 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 Z. Thus

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

and therefore u_Z is convex.

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

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

□

E Supplementary Material for Section 7

E.1 Verifying Stochastic Dominance in Two Dimensions

The goal of this section is to prove Lemma 5.

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 9. *Let α, β be positive finite Radon measures on $\mathbb{R}^n_{\geq 0}$ with $\alpha(\mathbb{R}^n_{\geq 0}) = \beta(\mathbb{R}^n_{\geq 0})$. A necessary and sufficient condition for $\alpha \succeq_1 \beta$ is that for all increasing*[30] *measurable sets A, $\alpha(A) \geq \beta(A)$.*

Proof of Claim 9: Without loss of generality assume that $\alpha(\mathbb{R}^n_{\geq 0}) = \beta(\mathbb{R}^n_{\geq 0}) = 1$.

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

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

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

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

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

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

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

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 $\int \tilde{f} d\beta > \int \tilde{f} d\alpha$. □

Due to Claim 9, to verify that a measure α stochastically dominates β in the first order, we must ensure that $\alpha(A) \geq \beta(A)$ for all increasing measurable sets A. This verification might still be difficult, since an increasing set can have fairly unconstrained structure. In Lemma 14 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 17. *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 18. *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 13. *Every k-discretized set S has only finitely many corners. Furthermore, $S = \cup_{c \in \mathcal{C}} B_c$, where $\mathcal{C}$ is the collection of corners of S.*

Proof of Lemma 13: 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,

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

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. □

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 14. *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 14: 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:

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

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

$$W_k = \left\{z \in \mathbb{R}^n_{\geq 0} : z_i > k \text{ for some } i\right\}; \qquad W^c_k = \left\{z \in \mathbb{R}^n_{\geq 0} : z_i \leq k \text{ for all } i\right\}.$$

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 $\lim_{k \to \infty} \int_{W^c_k} gdx = \int_{\mathbb{R}^n_{\geq 0}} gdx$, 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

$$\int_{A^u_k \cap W^c_k} gdx - \int_{A^l_k \cap W^c_k} gdx \leq |g|_{\sup} \left(V(A^u_k \cap W^c_k) - V(A^l_k \cap W^c_k) \right)$$

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 $z \in \mathbb{R}^n_{\geq 0}$, 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,

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

We thus have the bound

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

and therefore

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

In particular, we have

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

Since $\int_{A_k^u} g dx \ge \int_A g dx \ge \int_{A_k^l} g dx$, 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

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

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

We are now ready to prove Lemma 5.

Proof of Lemma 5:

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

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

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

Claim 10. *If $(a, w_2) \in R$, then $\zeta_a^b(w_2) \leq 0$.*

Proof of Claim 10: The inequality trivially holds unless there exists a $z_1 \in [a, b]$ such that $g(z_1, w_2) > h(z_1, w_2)$, so suppose such a z_1 exists. It must be that $(z_1, w_2) \notin R$, since both g and h are 0 in R. Indeed, because R is a decreasing set it is also true that $(\tilde{z}_1, w_2) \notin R$ for all $\tilde{z}_1 \geq z_1$. This implies by our assumption that

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

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

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

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

We now claim the following:

Claim 11. *Suppose that $\zeta_a^b(w_2^*) > 0$ for some $w_2^* \in [c_2, q_2)$. Then $\zeta_a^b(w_2) \geq 0$ for all $w_2 \in [w_2^*, q_2)$.*

Proof of Claim 11: Given that $\zeta_a^b(w_2^*) > 0$, our previous claim implies that $(a, w_2^*) \notin R$. Furthermore, since R is a decreasing set and $w_2 \geq w_2^*$, follows that $(a, w_2) \notin R$, and furthermore that $(c, w_2) \notin R$ for any $c \geq a$ in $[c_1, q_1)$. Therefore, we may write

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

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

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

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

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

as desired. □

We extend g and h to all of $\mathbb{R}^2_{\geq 0}$ by setting them to be 0 outside of $\mathcal{C}$. By Claim 14, to prove that $g \succeq_1 h$ it suffices to prove that $\int_A g dx dy \geq \int_A h dx dy$ for all sets A which are the union of finitely many bases. Since g and h are 0 outside of $\mathcal{C}$, it suffices to consider only bases $B_{z'}$ where $z' \in \mathcal{C}$, since otherwise we can either remove the base (if it is disjoint from $\mathcal{C}$) or can increase the coordinates of z' moving it to $\mathcal{C}$ without affecting the value of either integral.

We now complete the proof of Lemma 5 by induction on the number of bases in the union.

- **Base Case.**

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

$$\int_{B_r}(g-h)dxdy = \int_{r_2}^{q_2}\int_{r_1}^{q_1}(g-h)dz_1dz_2 = \int_{r_2}^{q_2}\zeta_{r_1}^{q_1}(z_2)dz_2.$$

  By Claim 11, we know that either $\zeta_{r_1}^{q_1}(z_2)\geq 0$ for all $z_2\geq r_2$, or $\zeta_{r_1}^{q_1}(z_2)\leq 0$ for all z_2 between p_2 and r_2. In the first case, the integral is clearly nonnegative, so we may assume that we are in the second case. We then have

$$\begin{aligned}\int_{r_2}^{q_2}\zeta_{r_1}^{q_1}(z_2)dz_2 \geq \int_{p_2}^{q_2}\zeta_{r_1}^{q_1}(z_2)dz_2 &= \int_{p_2}^{q_2}\int_{r_1}^{q_1}(g-h)dz_1dz_2\\ &= \int_{r_1}^{q_1}\int_{p_2}^{q_2}(g-h)dz_2dz_1.\end{aligned}$$

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

$$\int_{r_1}^{q_1}\int_{p_2}^{q_2}(g-h)dz_2dz_1 \geq \int_{p_1}^{q_1}\int_{p_2}^{q_2}(g-h)dz_2dz_1 = \int_{\mathcal{C}}(g-h)dxdy,$$

  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

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

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

  By Claim 11, we know that one of the two following cases must hold:

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

    In this case, we see that

$$\int_{z_2^{(k)}}^{z_2^{(k+1)}}\int_{z_1^{(k+1)}}^{z_1^{(k)}}(f-g)dz_1dz_2 = \int_{z_2^{(k)}}^{z_2^{(k+1)}}\zeta_{z_1^{(k+1)}}^{z_1^{(k)}}(w)dw \leq 0.$$

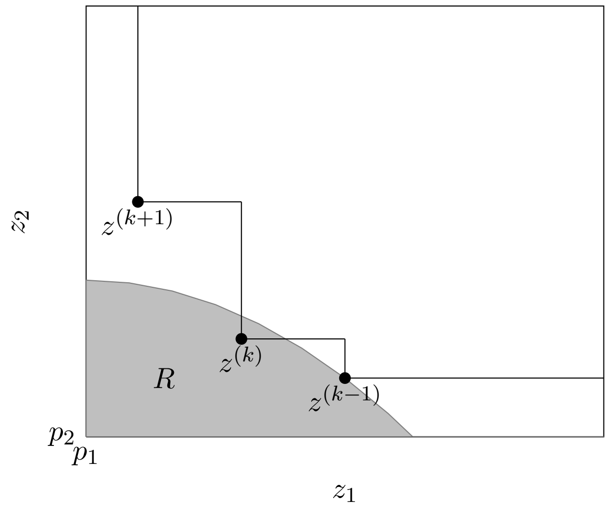


Figure 6: We show that either decreasing $z_2^{(k+1)}$ to $z_2^{(k)}$ or removing $z^{(k+1)}$ entirely decreases the value of $\int_A (f-g)$. In either case, we can apply our inductive hypothesis.

For notational purposes, we denote here by $(f-g)(S)$ the integral $\int_S (f-g)dz_1dz_2$ for any set S. We compute

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

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

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

In this case, we have

$$\int_{z_2^{(k+1)}}^{q_2} \int_{z_1^{(k+1)}}^{z_1^{(k)}} (f-g)dz_1dz_2 = \int_{z_2^{(k+1)}}^{q_2} \zeta_{z_1^{(k+1)}}^{z_1^{(k)}}(w)dw \geq 0.$$

Therefore, it follows that

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

where the final inequality follows from the inductive hypothesis.

□

E.2 Verifying Stochastic Dominance in Example 2

We sketch the application of Lemma 5 for verifying that $\mu_+|_{\mathcal{W}} \succeq_1 \mu_-|_{\mathcal{W}}$ in Example 2. We set $\mathcal{C} = [.16016, 1] \times [.09, 1]$ and $\mathcal{R} = Z \cap \mathcal{C}$, so that $\mathcal{W} = \mathcal{C} \setminus R$. We let g and h being the positive and negative parts of the density function of $\mu|_{\mathcal{W}}$, respectively, so that the density of $\mu|_{\mathcal{W}}$ is given by $g - h$. Since Z lies below *both* curves S_{top} and S_{right}, we know that integrating the density of μ along any horizontal or vertical line outwards starting anywhere on the boundary of Z yields a non-positive quantity, verifying the second condition of Lemma 5. In addition, on $\mathcal{W} = \mathcal{C} \setminus R$, 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 5, as $2/(1-z_1) + 3/(1-z_2) - 12$ is increasing. Finally, we verify the first condition of Lemma 5 by integrating $g - h$ over $\mathcal{C}$.

F Omitted Proofs from Section 8

F.1 Proof of Claim 7

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

- **Continuity.** Continuity of $\bar{\psi}$ follows from uniform continuity of ϕ and of $\|\cdot\|_1$.
- **Monotonicity.** Let $y \le y'$ coordinate-wise and let x be arbitrary. We must show that there exists an x' such that $\phi(x) - \|x - y\|_1 \le \phi(x') - \|x' - y'\|_1$. Set $x'_i = \max\{x_i, y'_i\}$. Since $x \le x'$, we have $\phi(x) \le \phi(x')$. We notice that if $x_i \ge y'_i$ then $x'_i = x_i$ and thus $|x'_i - y'_i| \le |x_i - y_i|$, while if $x_i \le y'_i$ then $|x'_i - y'_i| = 0$. Therefore, we have that $\|x - y\|_1 \ge \|x' - y'\|_1$ and thus $\phi(x) - \|x - y\|_1 \le \phi(x') - \|x' - y'\|_1$, 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

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

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

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

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

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

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

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

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

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

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

F.2 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 $u_1, u_2, \ldots \in \mathcal{U}(X) \cap \mathcal{L}_1(X)$ such that $\int_X u_i d\mu$ converges monotonically to the supremum value V, which we have proven is finite.[31] Since $\mu(X) = 0$, we may without loss of generality assume that $u_i(0^n) = 0$ for all u_i.

Pick an ordering $x_1, x_2, \ldots$ of the rational points in X. We note that $0 \leq u_j(x_i) \leq \|x_i\|_1$ for all points x_i, by the 1-Lipschitz condition. We now define a function u^* on the points x_i as follows:

- Set $u^*(0^n) = 0$ and initialize T to be the sequence $1, 2, 3, \ldots$.
- For $i = 1, 2, 3, \ldots$
  – Since $0 \leq u_j(x_i) \leq \|x_i\|_1$, there exists a subsequence T' of T such that $(u_j(x_i))_{j \in T'}$ converges.
  – Set $u^*(x_i) = \lim_{j \in T'} u_j(x_i)$ and replace T with T'.

It is straightforward to prove that the resulting function u^* is uniformly continuous on the points with rational coordinates. Therefore, we can uniquely extend u^* to a continuous function on all of X. It is furthermore simple to show that this function u^* indeed is increasing, convex, and 1-Lipschitz with respect to the ℓ_1 norm.

Finally, we argue that $\int_X u^* d\mu$ achieves the supremum value V. Let $\epsilon > 0$ be arbitrary. Since X is compact, there exists a finite collection of rational points $y_1, \ldots, y_k$ such that every point in X is within ϵ (in ℓ_1 distance) of some y_i. Denote by T^* the subsequence above immediately after $u^*(y_1), \ldots, y^*(y_k)$ have been defined.

By assumption, there exists an j^* such that for all $j > j^*$ it holds that $\int_X u_j d\mu$ is within an additive $\epsilon \cdot \mu_+(X)$ of the supremum value. Furthermore, since $y_1, \ldots, y_k$ is a finite collection of points, there exists $\tilde{j} > j^*$ such that $|u_j(y_i) - u^*(y_i)| < \epsilon$ for all $i = 1, \ldots, k$ and all $j > \tilde{j}$ in T^*.

[31] Finiteness is also obvious because X is bounded and the infimum problem is feasible.

For each $x \in X$, denote by $y_x \in X$ a y point which is closest to x in ℓ_1 distance. Since $\|x - y_x\|_1 < \epsilon$, we have $|u(x) - u(y_x)| < \epsilon$ for any feasible u. Therefore, we have

$$|u^*(x) - u_{\tilde{j}}(x)| \leq |u^*(x) - u^*(y_x)| + |u^*(y_x) - u_{\tilde{j}}(y_x)| + |u_{\tilde{j}}(y_x) - u_{\tilde{j}}(x)| < 3\epsilon.$$

Thus,

$$\begin{aligned}\int_X u^* d\mu &\geq \int_X (u_{\tilde{j}} - 3\epsilon) d\mu_+ - \int_X (u_{\tilde{j}} + 3\epsilon) d\mu_- \\ &= \int_X u_{\tilde{j}} d\mu - 6\epsilon\mu_+(X) \geq V - \epsilon\mu_+(X) - 6\epsilon\mu_+(X).\end{aligned}$$

Since ϵ was arbitrary and $\mu_+(X)$ is finite, we conclude that $\int u^* d\mu = V$.

G 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 Definition 5 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 μ. It is helpful therefore to consider only measures μ such that $\int \|x\|_1 d|\mu| < \infty$, to ensure that $\int u d\mu$ is finite for any utility function u. The measures in Examples 3 and 4 satisfy this property. There is a similar technical issue with our definition of convex dominance, in that the integrals in the definition might be infinite. 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 2 (weak duality) hold for unbounded μ, provided $\int \|x\|_1 d|\mu| < \infty$. Thus, tight certificates still certify optimality, even in the unbounded case. However, our strong duality proof relies on technical tools which require compact spaces, hence we do not know whether tight dual certificates are guaranteed to exist even when μ is unbounded.

To summarize our discussion so far, we can often transform measures and obtain an analogue of Theorem 1 for unbounded distributions (provided the distributions decay sufficiently quickly), and can easily obtain a weak duality result for such unbounded measures, but we do not know whether strong duality holds.

We finally discuss a technical issue with extending our two-item characterization of Theorem 5 to unbounded type spaces. The only real difficulty that arises with extending this result to unbounded type spaces is with the condition $\mu|_{\mathcal{W}} \succeq_2 0$. In our proof of Theorem 5, we obtain measures $\gamma_{\mathcal{W}}$ and θ and use Lemma 8 to deduce that $\mu|_{\mathcal{W}} \preceq_{cvx} \theta$, given the properties $\theta \preceq_2 \mu|_{\mathcal{W}}$ and $\int \|x\|_1 d\theta = \int \|x\|_1 d\mu|_{\mathcal{W}}$. Our current proof of Lemma 8, however, requires that θ and $\mu|_{\mathcal{W}}$ have bounded support. While we expect that this issue can be resolved, the current proof of Theorem 5 does not apply to unbounded type spaces. It is important to note, however, that if we use the weaker condition $\mu|_{\mathcal{W}} \succeq_1 0$, then the characterization indeed holds. (In fact, this is an equivalent way of stating the two-item characterization of [DDT13].) Indeed, the measures of Examples 3 and 4 satisfy this first-order dominance condition in region $\mathcal{W}$.

A similar issue arises with the second-order dominance constraint when extending our bundling theorem (Theorem 3) to unbounded type spaces. In addition, the "hard" direction of the proof (showing that the grand bundling conditions hold whenever bundling is optimal) requires technical lemmas which may not immediately apply without additional work.