UNIVERSITY *of York*

Discussion Papers in Economics

No. 23/02

Reduced-Form Allocations for Multiple Indivisible Objects under Constraints

Xu Lang
Zaifu Yang

Department of Economics and Related Studies
University of York
Heslington
York, YO10 5DD

Reduced-Form Allocations for Multiple Indivisible Objects under Constraints*

Xu Lang† and Zaifu Yang‡

Abstract: We examine the implementation of reduced-form allocation rules that assign multiple heterogeneous indivisible objects to many agents, with incomplete information and distributional constraints across objects and agents. To obtain implementability results, we adopt a lift-and-project approach, which enables us to find a general condition called total unimodularity, a well-recognized class of matrices with simple entries of -1, 0, or 1. This condition yields several new and general characterization results including those on hierarchies, bihierarchies, adjacency, and paramodularity. Our model and results extend and unify many well-known ones, cover both universal implementation and quotas-dependent implementation, and offer several new applications of practical interest.

JEL classification codes: D44, C65.

Keywords: Implementation, Reduced-form rules, Indivisible goods, Distributional constraints, Total unimodularity, Incomplete information.

*Date: April, 2023. This version is a revision of Lang and Yang (2021). We are very grateful to Jimmy Chan, Zhigang Chao, Eric van Damme, Dolf Talman, Charles Zheng, and many participants for their very helpful feedback on several occasions, including Chengdu Economic Theory Workshop July 2019 and Conferences on Mechanism and Institution Design June 2020 and July 2022. Any remaining errors are, however, our own responsibility.

†School of Economics, Southwestern University of Finance and Economics, Chengdu, China. Email: *langxu@swufe.edu.cn*

‡Department of Economics and Related Studies, University of York, York, UK. Email: *zaifu.yang@york.ac.uk*

1 Introduction

This paper aims to study a general implementation problem of reduced-form allocation rules for assigning multiple heterogeneous indivisible objects to many agents. The problem can accommodate a variety of distributional constraints across objects and agents in the environments of incomplete information. It has long been recognized that studying economies with indivisibility poses a serious challenge as indivisibility is an extreme form of nonconvexity; see e.g., Koopmans and Beckmann (1957), Debreu (1959), Arrow and Hahn (1971), Kelso and Crawford (1982), and Budish et al. (2013). We will identify a very general condition called total unimodularity for implementation and establish several general characterization results.

The reduced-form approach goes back to Myerson (1981) on auction design. In his paper, the seller's problem is expressed as a revenue maximization over the set of feasible and incentive compatible allocation rules and transfers. A buyer's incentive constraint is then used to express his transfers in terms of interim allocation probabilities, which reduce the problem to an optimization over interim allocation probabilities only, i.e., the *reduced-form*. To apply this approach, one should be able to describe the set of feasible interim allocation rules (or reduced forms). In single-item auctions, Maskin and Riley (1984) and Matthews (1984) first study the implementability condition, and Border (1991) derives a characterization, nowadays known as Border's theorem.

In this paper we investigate the reduced-form allocations of multiple indivisible items to many agents with a variety of distributional constraints, going beyond those traditional ones mentioned above. We briefly discuss several important cases of distributional constraints that our model intends to cover. In many practical situations, markets are regulated by distributional policies or constraints such that allocations across different objects are not independent. A typical example is the regional cap in Japanese residency matching program (Kamada and Kojima, 2015) that matches hospitals (agents) with doctors (objects). To regulate the geographical distribution of doctors, the total number of the doctors matched within a region is subject to a regional cap. Another important example is college admission. Higher education institutions usually set lower quotas for each of their particular areas of study. If the number of assigned students is less than this quota for a particular area, then the project has to be cancelled for that year (Biró et al., 2010, Ehlers et al., 2014, Fragiadakis and Troyan, 2016). Course allocation is yet another example (Budish and Cantillon, 2014, Budish et al., 2013, Sönmez and Ünver, 2010). In a university department, students seek to take multiple courses as part of their programs. Each student can take at most one seat in each course. For each course, there are ceiling and floor constraints on the number of seats. In addition, an aggregate capacity constraint may restrict the total number of seats for the courses within the department. In China, every region requires a certain minimum number of doctors as a floor constraint in its area hospitals and hospitals also face hierarchical constraints of recruiting doctors (Cheng and Yang, 2017). The presence of these distributional constraints poses a challenge for reduced-form allocation implementation: The problem cannot be treated as separate single-object problems and the reduced form for every agent is multidimensional.

To be precise, the problem under consideration concerns the allocation of a finite set of heterogeneous indivisible objects to a finite set of agents subject to distributional constraints. Objects can be private goods which will be consumed privately and independently by agents. Objects can be public or club goods like courses shared by students. Every agent may demand several objects and is associated with a finite set of types representing her private information about her preferences. There will be distributional constraints across different combinations of agents and objects. Constraints include floor and ceiling constraints across both agents and objects. We use a lift-and-project approach to obtain characterization results of implementability. This approach was first introduced in polyhedral combinatorics (Balas, 2001, Balas and Pulleyblank, 1983) and later used by Vohra (2013) to study linear characterizations for combinatorial objects, including reduced-form auctions. Briefly speaking, in this approach, by lifting, the combinatorial object of interest (i.e., reduced forms) is first formulated by a linear system in some higher-dimensional space. Then by constructing a projection cone and finding its finite generators, it gives arise to the linear system of interest. We find this approach surprisingly powerful.

To obtain a complete description of the generators, we investigate geometric and combinatorial properties of the projection cone. We discover a general sufficient condition on the projection cone such that a complete description of the generators is possible. The general condition underlying our characterization is called *total unimodularity*, which is probably the most general one we could possibly have. It concerns so-called totally unimodular matrices, a class of well-behaved and well-studied matrices with simple entries of -1, 0, or 1 in discrete optimization (Schrijver, 1986). It is also well-known that there exists a polynomial time recognition algorithm for totally unimodular matrices; see Seymour (1980, 1981) and Schrijver (1986). In other words, it is very easy to verify the general condition. We show that if the projection cone preserves total unimodularity, then the generators of the projection cone can be completely described.

We identify four large classes of constraint structures where the projection cone preserves total unimodularity: *Hierarchies, bihierarchies, adjacency, and paramodularity.* Hierarchical structures are common in organizations (firm, hospital, or university) and various markets, and have been well studied in the literature; see e.g., Williamson (1975) and Demange (2004). They are called laminars in mathematics; see Fujishige (2005). Bihierarchies are the union of two disjoint hierarchies as a generalization of hierarchies and are recently investigated by Budish et al. (2013). Adjacency is a basic notion from networks and graphs that reflects close relationships between agents or objects; see e.g., Bondy and Murty (1976) and Schrijver (1986). Adjacent agents can cooperate and obey shared rules and conventions. The property of adjacency has been used for cooperative games with communication structure by e.g., Myerson (1977) and Herings et al. (2010). Paramodularity has been used by Che et al. (2013) and others and is closely related to the basic concept of submodularity/supermodularity in combinatorial optimization; see e.g., Fujishige (2005). It is imposed upon the floor and ceiling constraints of the supply side. Roughly speaking, submodularity means substitutability.

Our framework allows us to deal with two major classes of implementation problems: Universal or quotas-independent implementation and quotas-dependent implementation. Universal implementation does not depend on the specification of quotas, while quotas-dependent implementation relies on the specification of quotas. In many situations, the designer may have no information on capacity and a universal implementation is therefore very desirable (Budish et al., 2013), while other situations may face capacity constraints and therefore quotas-dependent implementation may arise (Che et al., 2013). We offer characterization results on both universal implementation and quotas-dependent implementation. Our characterization results are algebraic and very general, covering a variety of distributional constraints such as floor constraints, ceiling constraints, and mixed floor and ceiling constraints.

Our first major result (Theorem 1) establishes a general characterization on universal implementation, showing that *total unimodularity is a sufficient condition*. Our second and third major results (Theorems 2 and 3) prove that *hierarchies, bihierarchies, and adjacency* each suffice to guarantee total unimodularity. Our fourth result (Theorem 4) concerns a general characterization on quotas-dependent implementation where capacity constraints are paramodular and a model of multiple heterogeneous goods with multi units for each good is considered. This generalizes a well-known result of Che et al. (2013) on a multi-unit model of a single good and is similar to a recent independent result of Zheng (2021) which appeared earlier than ours Theorem 4. Our approach is quite different from those given by Che et al. (2013) and Zheng (2021).[1] It is worth pointing out that both total unimodularity and the lift-and-project approach are used to establish our characterization results on both universal implementation and quotas-dependent implementation. We also discuss a variety of applications with practical interest including systems of multiple suppliers, assignment markets with both ordinal and cardinal utilities, allocation of club goods, assignment of radio spectrum licenses, bilateral trade, and compromise model.

We close this introductory section by further discussing some literature. For an auction model with multiple identical goods and capacity constraints, Che et al. (2013) develop a network flow method to characterize the implementability condition. In their method, the implementation problem is first transformed into a feasible network flow problem and then existence results from graph theory are invoked to obtain their characterization results. Goeree and Kushnir (2016, 2022) propose a geometric approach (i.e., support function of convex sets) to study implementation in social choice problems. Alternative characterizations have also been found. Manelli and Vincent (2010) and Gershkov et al. (2013) establish an important equivalence of Bayesian and dominant strategy implementation. Hart and Reny (2015) obtain a characterization by majorization. Alaei et al. (2019) study a polymatroidal decomposition method and obtain their characterization results. Meanwhile, Gopalan et al. (2015), Cai et al. (2012), and Alaei et al. (2019) examine the computational complexity of the reduced-form approach. A common feature of these models is that there are no side constraints across different items.

[1]More precisely, Che et al. (2013) use a network flow approach and Zheng (2021) applies a separating hyperplane approach.

The rest of this paper is organized as follows. Section 2 presents a general model on reduced-form implementation. Section 3 introduces our main characterization theorems on implementation. Section 4 discusses the lift-and-project approach and several lemmas which play a key role in proving the characterization theorems. Section 5 discusses several applications of practical interest. Section 6 concludes.

2 The Model

We study a model in which a finite set A of m types of heterogeneous indivisible objects (e.g., workers, doctors, goods, and courses) is allocated to a finite set N of n different agents (e.g., firms, hospitals, and students) under a variety of constraints. Each object in the set A can accommodate multiple identical units. A **pure outcome** $x = (x(i, j))$ is described as an $n \times m$ matrix indexed by all agents i and objects j, where each entry $x(i, j) \in \mathbb{Z}_+$ is the quantity of object j that agent i receives.[2] Note that it is possible for each agent to receive several objects with different units and also for one object like course to be shared by several agents. Generally speaking, in the allocation of private goods, no unit of any object will be assigned to more than one agent, while in the case of shared or public goods, every object can be jointly consumed by multiple agents.

A set $G \subseteq N \times A$ of agent-object pairs is called a **constraint set**. Every pair (i, j) is called a **singleton**. For a constraint set G, we define $x(G) = \sum_{(i,j)\in G} x(i, j)$. Each constraint set G is associated with two integer numbers $b(G), c(G) \in \mathbb{Z}_+$ with $b(G) \leq c(G)$ as its floor and ceiling **quotas**, respectively. We assume $b(\emptyset) = c(\emptyset) = 0$. A collection of constraint sets $\mathcal{G} \subseteq 2^{N\times A}$ is called a **constraint structure**. A constraint structure $\mathcal{G}$ and a quotas system $b, c : \mathcal{G} \to \mathbb{Z}_+$ define a system $(\mathcal{G}, b, c)$, which restricts the set of feasible outcomes. We say a pure outcome x is **feasible** if

$$b(G) \leq x(G) \leq c(G) \;\text{ for each }\; G \in \mathcal{G}. \tag{1}$$

Clearly, for every feasible pure outcome x, we have $x(i, j) \geq 0$ for all $(i, j) \in N \times A$ and so $\mathcal{G}$ contains all singletons.

Let X denote the set of feasible pure outcomes. A **random outcome** is a matrix

[2] The sets $\mathbb{Z}$ and $\mathbb{Z}_+$ stand for the sets of all integers and all nonnegative integers, respectively. Similarly, we can define the sets $\mathbb{Z}^m$ and $\mathbb{Z}^m_+$ for the m-dimensional integral vectors.

$x = (x(i,j))$ indexed by agents and objects where $x(i,j) \geq 0$ is a fractional allocation of object $j \in A$ assigned to agent $i \in N$. A random outcome x is **feasible** if it can be described as a lottery over the set of feasible pure outcomes, that is, if there exist nonnegative numbers λ^k summing up to one and feasible pure outcomes $x^k \in X$ such that

$$x = \sum_{x^k \in X} \lambda^k x^k. \tag{2}$$

Let $\Delta(X)$ denote the set of all feasible random outcomes. We let $(N, A, \mathcal{G}, b, c)$ denote an allocation problem.

Our model covers a variety of allocation problems. We briefly discuss two major allocation problems and their implications on their quotas system. Suppose each object j has a fixed supply $z(j) \in \mathbb{Z}_+$ and we denote the supply vector $z = (z(j)) \in \mathbb{Z}_+^m$ for all objects. In any allocation problem of private goods, usually no unit of any object $j \in A$ will be assigned to more than one agent. Then, for each object j, the total demand $x(N \times \{j\})$ is bounded above by $z(j)$ and we may assume $N \times \{j\} \in \mathcal{G}$. It also implies $x(G(j)) \leq c(G(j)) = z(j)$ where $G(j) = \{(i,j) \mid (i,j) \in G\}$ for $G \in \mathcal{G}$. However, for any allocation problem of shared goods like courses or public goods, every unit of any object $j \in A$ can be typically shared by multiple agents and we may have $x(G(j)) \leq c(G(j))$ where $c(G(j))$ is no less than $z(j)$.

Our model is flexible to incorporate two large classes of constraints: **quotas-independent** constraints and **quotas-dependent** constraints. To illustrate their difference, we write the feasibility constraints in the following matrix form

$$\begin{bmatrix} M \\ -M \end{bmatrix} x \leq \begin{bmatrix} c \\ -b \end{bmatrix}, \tag{3}$$

where M is the constraint matrix for constraint structure $\mathcal{G}$. The quotas-independent problems impose constraints on the left hand side of the system (i.e., on $\mathcal{G}$) only. The quotas-dependent problems impose constraints on the right hand side of the system (i.e., on (b,c)) only. The two classes of constraints then correspond to two different notions of implementation problems that we will discuss later. In particular, we call *c the supply vector.*

Generalized capacity function. We now consider a generalized capacity function which extends the domain of quotas system (b,c) from $\mathcal{G}$ to the collection of all subsets of $N \times A$ and provides a unified representation for the ceiling and floor constraints. Define the set of all ordered pairs of disjoint constraint sets $3^{N\times A} = \{(U,V) : U,V \subseteq N \times A,\ U \cap V = \emptyset\}$. We introduce the concept of the generalized capacity function.

Definition 1. *Let $(N, A, \mathcal{G}, b, c)$ be given. The generalized capacity function is a biset function $\beta : 3^{N\times A} \to \mathbb{R}$ where for every $(U,V) \in 3^{N\times A}$,*

$$\beta(U,V) = \max\{x(U) - x(V) \mid x \in \Delta(X)\}. \tag{4}$$

Intuitively, the generalized capacity function gives all supporting hyperplanes of the set of feasible outcomes with $-1, 0, 1$-valued weights. The bounds defined by β are effective, i.e., for each U and V, the bounds $\beta(U, \emptyset)$ and $\beta(\emptyset, V)$ can be attained by some feasible outcomes. When condition (1) completely describes the set of feasible random outcomes (i.e., a bihierarchy), a supporting hyperplane argument implies that $P = \{x \mid x(U) - x(V) \leq \beta(U,V) \text{ for all } (U,V) \in 3^{N\times A}\}$ also describes the original feasible set, since it contains inequalities of the constraint sets in $\mathcal{G}$ as a subset.

Note that from the definition of a supporting hyperplane, β is **sublinear**: for all $(U,V) \in 3^{N\times A}$,

$$\beta(U,V) \leq \beta(U,\emptyset) + \beta(\emptyset,V). \tag{5}$$

We say β is **linear** if the inequality (5) holds with equality for all $(U,V) \in 3^{N\times A}$. In Section 3, we will discuss the implication of this property for our characterization (Corollary 1).

2.1 Quotas-independent constraints

The first class of constraints allows an unspecified structure of quotas. A well-known example concerns the classical model of assignment markets; see Koopmans and Beckmann (1957), Shapley and Shubik (1971), Crawford and Knoer (1981), Demange et al. (1986), Bogomolnaia and Moulin (2001). Note that in the classical model, every buyer demands one item and every seller supplies one item, i.e., the supply vector $c = (1, \cdots, 1)$ is the vector of ones. Here we actually allow every buyer to demand more than one item and the supply vector c to be any nonnegative integer vector. Following Budish et al. (2013), we call a constraint structure $\mathcal{G}$ a **canonical two-sided constraint structure**, if $\mathcal{G}$ contains all sets $\{i\} \times A$ for each $i \in N$ (i.e., all rows) and all sets $N \times \{j\}$ for each $j \in A$ (i.e., all columns). This is a special and important class of constraint structures obviously covered by the general framework of constraint structures described above.

We introduce three large classes of constraint structures on which several characterization results will be built. The first one is the class of hierarchies, which has been studied by Williamson (1975) and Demange (2004) in different contexts, also called laminars in mathematics (Fujishige, 2005).

Definition 2. *A constraint structure $\mathcal{G}$ is a **hierarchy** (or a laminar) if for all $G, G' \in \mathcal{G}$,*

$$G \subset G', \ \ or \ G' \subset G, \ \ or \ G' \cap G = \emptyset.$$

Our second one concerns a richer and more general class of constraint structures due to Budish et al. (2013), called bihierarchies. The canonical two-sided assignments are included as a special case.

Definition 3. *A constraint structure $\mathcal{G}$ is a **bihierarchy** if it is the union of two disjoint hierarchies $\mathcal{G}_1$ and $\mathcal{G}_2$, i.e., $\mathcal{G} = \mathcal{G}_1 \cup \mathcal{G}_2$ and $\mathcal{G}_1 \cap \mathcal{G}_2 = \emptyset$.*

For bihierarchical constraint structures, Budish et al. (2013) obtain a generalization of Birkhoff-von Neumann theorem which is used to characterize their feasible random outcomes; see their Theorem 1.[3] For our current model, we will use bihierarchies as an example of our general conditions to obtain implementation results.

We now introduce the third class of constraint structures, called adjacency-a basic concept used in networks and graphs (Bondy and Murty (1976) and Schrijver (1986)). The constraint structure will be described by the structure of a graph $H = (N, E)$, where N is the set of agents which are represented by the vertices of the graph, and E is the set of edges which indicate the relationship between agents. An edge is denoted by a pair $e = \{i_1, i_2\}$ of vertices $i_1, i_2 \in N$. Two agents or vertices i_1 and i_2 are said to be *adjacent* if they are connected by edge $\{i_1, i_2\}$. Adjacent agents have some kind of relationship and can cooperate and obey shared rules or conventions (i.e., constraints)

[3]Note that the model and objective of Budish et al. (2013) are considerably different from ours. They develop ex ante efficient and fair random allocation mechanisms for environments with various constraints. We deal with a general model with incomplete information and various constraints and investigate interim incentive-compatible allocation mechanisms. In our model every agent $i \in N$ has a finite set T_i of types, while in Budish et al. (2013) incomplete information is not discussed and in their model the set T_i of every agent i could be understood to contain only one element.

to advance their interest; see e.g., Myerson (1977) and Herings et al. (2010) for the use of graphs in cooperative games. The constraint structure $\mathcal{G}$ is the set E of all edges in the graph H. For this constraint structure, we focus on the basic case of multiple identical goods or services and therefore ignore the set A of objects by using only the sets N and E.

Definition 4. *A constraint structure $\mathcal{G}$ with the set N of agents and the set A of identical objects has* ***the adjacency property*** *if $\mathcal{G}$ is given by the set of all edges in the graph $H = (N, E)$, as described above, i.e., $\mathcal{G} = E$.*

In the current setting, only adjacent agents will face their relevant constraints. In Section 5, we will discuss this notion of adjacency in the allocation of bands of radio spectrum (Rothkopf et al., 1998). Adjacency is related to but different from the notion of consecutiveness; see Greenberg and Weber (1986).

Below we give three examples of different constraint structures. The first two concern bihierarchies and the last one concerns adjacency.

Example 1 (Buyer-seller problem)**.** Suppose that there are two buyers $N = \{i_1, i_2\}$ and two sellers $A = \{j_1, j_2\}$. Each buyer i has at most $c(i) \in \mathbb{Z}_+$ units of demand and each seller j has at most $c(j) \in \mathbb{Z}_+$ units of supply. The following system of linear inequalities describes all possible outcomes:

$$\begin{array}{rl} \sum_{j\in A} x(i,j) \leq c(i) & \text{for every } \ i \in N, \\ \sum_{i\in N} x(i,j) \leq c(j) & \text{for every } \ j \in A, \\ x(i,j) \in \mathbb{Z}_+ & \text{for every } \ i, j. \end{array}$$

Example 2 (Course allocation problem)**.** Suppose that there are two students $N = \{i_1, i_2\}$, one compulsory courses c_1, and three optional courses $O = \{o_1, o_2, o_3\}$. Every optional course $j \in O$ faces a floor constraint which requires the course being selected by at least one student for it to open and has at most two seats; Every student i is required to take at least one and at most two of the optional courses. The family of constraint sets is given by (i, o) for every $i \in N$ and $j \in O \cup \{c_1\}$, $\{i\} \times O$ for every $i \in N$, and $N \times \{o\}$ for every $j \in O \cup \{c_1\}$, which is a bihierarchy, having a canonical two-sided constraint structure. The following system of linear inequalities describes all possible outcomes:

$$\begin{aligned} 1 \le \textstyle\sum_{j\in O} x(i,j) \le 2 & \quad \text{for every } \; i \in N, \\ 1 \le \textstyle\sum_{i\in N} x(i,j) \le 2 & \quad \text{for every } \; j \in O, \\ \textstyle\sum_{i\in N} x(i,c_1) = 2 & \\ x(i,j) \in \{0,1\} & \quad \text{for every } \; i,j. \end{aligned}$$

This following example demonstrates that adjacency and bihierarchy do not imply each other and are therefore two independent concepts.

Example 3 (Adjacency)**.** Assume that we have five agents denoted by $N = \{1,2,3,4,5\}$ and many identical objects. The graph is given by $H = (N,E)$ with $E = \{\{1,4\},\{1,5\},\{2,4\},\{2,5\},\{3,4\},\{3,5\}\}$. Each pair $\{i,j\} \in E$ has a floor quota $b(\{i,j\})$ and a ceiling quota $c(\{i,j\})$. The constraint structure $\mathcal{G}$ is given by the set E of edges. The following system of linear inequalities describes all possible outcomes:

$$\begin{aligned} b(\{i,j\}) \le x(i) + x(j) \le c(\{i,j\}) & \quad \text{for every } \; \{i,j\} \in E, \\ x(i), x(j) \in \{0,1\} & \quad \text{for every } \; \{i,j\} \in E. \end{aligned}$$

Observe that $\mathcal{G}$ has the adjacency property but is not bihierarchical as it contains three distinctive hierarchies $\mathcal{G}_1 = \{\{1,4\},\{2,5\}\}$, $\mathcal{G}_2 = \{\{1,5\},\{3,4\}\}$, and $\mathcal{G}_3 = \{\{2,4\},\{3,5\}\}$.

On the other hand, if we have the bihierarchical constraint structure

$$\mathcal{G}' = \{\{1,2\},\{1,2,3\},\{3,4,5\}\},$$

this obviously does not satisfy the adjacency property.

2.2 Quotas-dependent constraints

The second class of constraints requires detailed information about the quotas. Suppose the constraint structure is unrestricted and contains all possible sets of agent-object pairs:

$$b(G) \le x(G) \le c(G), \text{ for each } G \subseteq N \times A. \tag{6}$$

We introduce the following family of paramodular constraints imposed upon the supply side.

Definition 5. *The pair* $(b, c) : 2^{N\times A} \to \mathbb{R}_+$ *is* ***paramodular*** *if*

(1) $-b$ *and* c *are submodular.*[4]

(2) b *and* c *are compliant: for all* G, G', $c(G) - b(G') \geq c(G \setminus G') - b(G' \setminus G)$.

Submodularity on the ceiling capacity roughly says that marginal supply of goods over agents and types of goods is decreasing, while supermodularity on the floor capacity says that marginal supply of goods over agents and types of goods is increasing. Compliance requires consistency across the upper and lower bounds. Note that this definition has been used by Che et al. (2013) for the case of multiple units of a single goods, in which both c and b are defined on each group of agents. In the current setting of heterogeneous goods, both c and b involve each group of both agents and types of goods. It is well known that if the pair (b, c) is paramodular, then the upper and lower bounds determined by the generalized capacity function β are also paramodular and coincide with (b, c): $\beta(G, \emptyset) = c(G)$ and $\beta(\emptyset, G) = b(G)$ for all $G \subseteq N \times A$.

Paramodularity has been widely used in the literature and encompasses a broad range of problems, including hierarchies and matroids as special cases. The following example shows that bihierarchy fails paramodularity. It is a simple canonical bihierarchy (i.e., a simple assignment market model). Its upper quotas are not submodular and hence the constraints are not paramodular.

Example 4. *Let* $N = \{i_1, i_2\}$ *be two buyers and let* $A = \{j_1, j_2\}$ *be two sellers. Suppose the constraint structure is canonical (with row and column constraints) with quotas* $c(\{(i, j_1), (i, j_2)\}) = 1$, $c(\{(i_1, j), (i_2, j)\}) = 1$, $c(\{(i, j)\}) = 1$, *and* $b = 0$. *Then for all* $i \in N$ *and* $j \in A$,

$$x(i, j_1) + x(i, j_2) \leq 1,$$
$$x(i_1, j) + x(i_2, j) \leq 1,$$
$$0 \leq x(i, j) \leq 1.$$

It follows that the upper bound for $G = \{(i_1, j_1), (i_1, j_2), (i_2, j_2)\}$ *is found by solving*

[4] Let E be a finite ground set. We say $f : 2^E \to \mathbb{R}_+$ is submodular if for all $G, G' \subseteq E$, $f(G) + f(G') \geq f(G \cap G') + f(G \cup G')$

the problem $\max_{x \in \Delta(X)} x(G)$*, which is equal to* 2*. Then*

$$c(\{(i_1, j_1), (i_1, j_2)\}) + c(\{(i_1, j_2), (i_2, j_2)\}) = 2 < 3 = c(G) + c(\{(i_1, j_2)\}).$$

Hence c *fails to be submodular.*

2.3 Reduced-form implementation

Given any welfare objective at the ex ante or interim stage, we are interested in Bayesian incentive compatible allocation rules that may allow random outcomes. Using the reduced-form approach, we first optimize over the set of incentive compatible interim allocation rules and find an optimal solution. We then ask whether this interim optimal solution can be implemented by an ex post feasible allocation rule or not. The reduced-form implementation problem is to characterize implementable allocation rules based on interim incentive compatibility.

We first extend $(N, A, \mathcal{G}, b, c)$ to the incomplete information environment. Every agent $i \in N$ is associated with a finite set T_i of possible types, which represents agent i's private information about her preference. A type $t_i \in T_i$ may represent agent i's preference ordering $\succsim_i$ over the set A of objects, or it may determine her cardinal utility function, i.e., a payoff vector $v_i \in \mathbb{R}^A$ that assigns a valuation $v_i(j)$ for object j. Hence our model allows for domains that cover ordinal and cardinal preferences. Let $T = \times_{i \in N} T_i$ denote the entire type set, i.e., the product of the type set T_i over all agents $i \in N$, and $T_{-i} = \times_{j \neq i} T_j$. For every $i \in N$, let $\lambda^i : T_i \to \Delta(T_{-i})$ be a belief function, i.e., $\lambda^i(t_{-i}|t_i)$ is the probability that agent i assigns to other agents' type $t_{-i} \in T_{-i}$ when i's type is t_i. We assume that there exists a common prior probability $\lambda \in \Delta(T)$ such that the beliefs of the agents are the posteriors, and $\lambda(t) > 0$ for all $t \in T$. Let λ_i denote agent i's marginal probability of λ.

A **feasible ex post allocation rule** $p : T \to \mathbb{R}_+^{N \times A}$ assigns a feasible random outcome $p(\cdot, t)$ for each type profile $t \in T$, where $p(i, j, t)$ is a fractional quantity of object $j \in A$ assigned to agent $i \in N$. In particular, an allocation rule is deterministic if it assigns a pure outcome for each type profile. A feasible ex post allocation rule p induces an interim allocation rule $Q = (Q_i)_{i \in N}$, where $Q_i : T_i \to \mathbb{R}^A$ is agent i's interim expected random allocation. For each $i \in N$, $t_i \in T_i$, and $j \in A$,

$$Q_i(t_i, j) := \sum_{t_{-i} \in T_{-i}} p(i, j, t)\lambda^i(t_{-i}|t_i). \tag{7}$$

An interim allocation rule Q is **implementable** if there exists a feasible ex post allocation rule p such that (p, Q) satisfies (7). We then say Q is the **reduced form** of p and p implements Q. Let $\mathcal{Q}$ denote the set of all implementable interim allocation rules. Let $(N, A, \mathcal{G}, T, \lambda, b, c)$ represent **the implementation problem**. We call $(N, A, \mathcal{G}, T)$ the **implementation structure**. Define $d := |N||A||T|$ and $l := \sum_{i \in N} |T_i||A|$ and $r := |\mathcal{G}||T|$. Note that the set of feasible ex post allocation rules is defined by a set of linear inequalities and the reduced form operation is a linear map. Hence for the given set $\mathcal{Q}$ the implementable interim rules can be also defined by a set of linear inequalities:

$$\mathcal{Q} = \{Q \in \mathbb{R}^l \mid MQ \leq u\}$$

for some matrix M and vector u. The goal of our implementation problem is to find a linear system of (M, u) that describes the set $\mathcal{Q}$. The system (M, u) is called *a linear characterization* on the set of implementable interim allocation rules.

2.4 Implementability conditions

We first present two intuitive necessary conditions for implementability and discuss their sufficiency in the next sections. For any $S_i \subseteq T_i \times A$, $i \in N$, let $S = \cup_{i \in N} S_i$. For any $t \in T$ and $S \subseteq \cup_{i \in N}(T_i \times A)$, we define their intersection by

$$I(t, S) = \{(i, j) \in N \times A \mid (t_i, j) \in S_i\}.$$

Proposition 1. (Condition A) *Let $(N, A, \mathcal{G}, T, \lambda, b, c)$ be an implementation problem. If Q is implementable, then for all $S_i^+, S_i^- \subseteq T_i \times A$, $S_i^+ \cap S_i^- = \emptyset$, for each $i \in N$,*

$$\sum_{i \in N} [\sum_{(t_i, j) \in S_i^+} Q_i(t_i, j)\lambda_i(t_i) - \sum_{(t_i, j) \in S_i^-} Q_i(t_i, j)\lambda_i(t_i)] \\ \leq \sum_{t \in T} \lambda(t)\beta(I(t, S^+), I(t, S^-)). \tag{A}$$

Condition (A) provides a compact description of the implementability condition with both the ceiling and floor constraints, as the left hand side may contain both positive and negative entries and the generalized capacity β is implicitly determined by the ceiling and floor in condition (4). In the Appendix, we provide an example to illustrate Condition (A).

We provide an intuitive interpretation for Condition (A). Consider a market where each agent i can be a seller or a buyer of each good, and (S_i^+, S_i^-) defines a trading strategy of agent i. That is, agent i buys one unit of good j if her type is in $S_i^+ \cap (T_i \times \{j\})$ and sell one unit of good j if her type is in $S_i^- \cap (T_i \times \{j\})$. Given a strategy profile of the agents, Condition (A) requires that the ex ante expected demand and supply over all agents and goods are approximately equal, i.e., the difference between the demand and supply are bounded by the maximum difference constrained by the generalized capacity function. The feasibility requires that the condition holds for all possible strategy profiles of agents.

The following proposition provides a subset of inequalities in Condition (A), which contains only separate inequalities for ceiling and floor constraints by setting $S^- = \emptyset$ or $S^+ = \emptyset$. Hence it also gives a necessary condition for implementability.

Proposition 2. (Condition B) *Let $(N, A, \mathcal{G}, T, \lambda, b, c)$ be an implementation problem. If Q is implementable, then for all $S_i \subseteq T_i \times A$, for each $i \in N$,*

$$-\sum_{t\in T}\lambda(t)\beta(\emptyset, I(t,S)) \leq \sum_{i\in N}\sum_{(t_i,j)\in S_i} Q_i(t_i,j)\lambda_i(t_i) \leq \sum_{t\in T}\lambda(t)\beta(I(t,S),\emptyset). \qquad \text{(B)}$$

It is well known that when there is a single type of good with multiple units, Condition (B) reduces to the characterization condition obtained by Che et al. (2013). Condition (B) further reduces to the classical condition of Border (1991) for a single object. In Section 3, we will provide sufficient conditions on an implementation problem such that Condition (A) or (B) is necessary and sufficient for implementability.

3 Main Results

In this section, we establish our main characterization theorems for the implementation problem. To obtain these results, we use a lift-and-project approach. This approach starts with a linear system in terms of both ex post and interim allocation rules. By projecting away the variables of ex post allocation rules, we obtain a linear system of interim allocation rules. The procedure then reduces the implementation problem to a problem of characterizing the generators of the projection cone. Since the analysis is involved, we first state our characterization theorems in this section and then discuss the lift-and-project approach in the next section. Most of the proofs will be deferred to the Appendix.

For the convenience of the reader we first review several basic mathematical concepts concerning cones, generators, and extreme rays that will be often used in our analysis. In the paper, 0 can be the number of 0, a vector of 0's, or a matrix of 0's whose dimension can be understood from the context. Let $x_1, \ldots, x_k \in \mathbb{R}^q$ be given vectors. A linear combination $\alpha_1 x_1 + \cdots + \alpha_k x_k$ is *conic* if $\alpha_1, \ldots, \alpha_k \geq 0$. The *cone* generated by a finite set $X \subset \mathbb{R}^q$ is the set of all conic combinations of the elements from X, denoted by $\text{cone}(X)$, and we call the vectors in X the *generators*. A cone P is polyhedral if $P = \{x \in \mathbb{R}^q \mid Mx \leq 0\}$ for some matrix M. We say P is *pointed* if $Mx = 0$ implies $x = 0$. Otherwise, we say P is *non-pointed*. A nonzero element x of a pointed cone P is called *an extreme ray* if there are $q-1$ linearly independent constraints binding at x. For detail we refer to Schrijver (1986). In the following, we often say $\{0,1\}$ or $\{-1,0,1\}$ rays or generators if all of their entries are in $\{0,1\}$ or $\{-1,0,1\}$.

We first consider the implementation problem $(N, A, \mathcal{G}, T, \lambda, b, c)$ with general constraint structures and quotas. To proceed, we introduce the following two $\{0,1\}$ matrices B and C which will be central for our analysis.

- For every $p \in \mathbb{R}^d$, $Q \in \mathbb{R}^l$ and $b, c \in \mathbb{R}^r$, we define the corresponding probability weighted variables x, y, b^λ, and c^λ by multiplying each $p(i,j,t), b(G), c(G)$ by $\lambda(t)$, and each $Q_i(t_i, j)$ by $\lambda_i(t_i)$.

- Matrix C: An $r \times d$ incidence matrix where each row is indexed by (G, t), each column is indexed by (i, j, t), and the entry in row (G, t) and column (i, j, t') is 1 if $(i, j) \in G$ and $t = t'$, and 0 otherwise.

- Matrix B: An $l \times d$ incidence matrix where each row is indexed by (i, t_i, j), each column is indexed by (i, j, t), and the entry in row (i, t_i, j) and column (i', j', t') is 1 if $i = i'$, $j = j'$, and $t_i = t'_i$, and 0 otherwise.

We use Example 1 to illustrate how the matrices B and C are constructed. For any positive integer k, let I_k denote the identity matrix of order k.

Example 5. (Example 1 continued) *Suppose there are two agents, two goods, and each agent has two types. The constraint matrices B and C are given by* $B = \begin{bmatrix} B_1 & \\ & B_2 \end{bmatrix}$, $C = \begin{bmatrix} C_{1,1} & C_{1,2} \\ C_{2,1} & \\ & C_{2,2} \\ & I_{16} \end{bmatrix}$, *where* $B_1 = B_2 = [b_{k,l}]_{4\times 8}$ *with* $b_{k,l} = 1$ *if* $k = (i, t_i, j)$ *and* $l = (i, j, t)$, *and* $b_{k,l} = 0$ *otherwise, and* $C_{1,1} = C_{1,2} = I_8, C_{2,1} = C_{2,2} = \begin{bmatrix} I_4 & I_4 \end{bmatrix}$. *Here each B_k denotes the constraint matrix of the reduced-form equalities for each good j_k, and C_1 and C_2 denote the constraint matrices of the canonical row and column constraints, and I_{16} denotes the constraint matrix for singletons.*

The implementation system associated with the matrices B and C arising from the implementation problem described above is given by

$$F = \{(x, y) \in \mathbb{R}^d \times \mathbb{R}^l \mid b^\lambda \leq Cx \leq c^\lambda, y = Bx\}. \tag{8}$$

We define the **projection cone** of the set F by

$$W = \{(f, g, h) \in \mathbb{R}^l \times \mathbb{R}^r \times \mathbb{R}^r \mid -B^\top f + C^\top g - C^\top h = 0, g \geq 0, h \geq 0\}. \tag{9}$$

The constraint matrix of the projection cone W is given by

$$M(W) = \begin{bmatrix} -B^\top & C^\top & -C^\top \\ 0 & I_r & 0 \\ 0 & 0 & I_r \end{bmatrix}. \tag{10}$$

Remark 1. The projection cone W depends on the implementation structure $(N, A, \mathcal{G}, T)$ but not on the quotas and the beliefs. It implies that the same set of generators would arise if two problems differ only in the quotas and beliefs.

The following lemma provides a conic characterization of implementability with all vectors in the projection cone and is the common starting point for all of our following characterizations.

Lemma 1. *(Conic characterization) $y \in \mathbb{R}^l$ is implementable if and only if*

$$f^\top y \leq g^\top c^\lambda - h^\top b^\lambda \quad \text{for all } (f, g, h) \in W. \tag{11}$$

Overview. From now on, we will discuss two major problems of implementation. One is the **universal implementation** for which the constraint structure is quotas-independent and imposed on the demand side (i.e., W), while the other is the **quotas-dependent implementation** for which the constraint structure is quotas-dependent and imposed on the supply side (i.e., (b, c)). For each of these two problems, we provide sufficient conditions such that a characterization with a finite set of inequalities is possible.

(1) *Universal implementation.* Given certain conditions on the implementation structure $(N, A, \mathcal{G}, T)$ and hence the projection cone W, we show that it is possible to obtain an explicit characterization of the generators of W in Lemma 1 (Theorems 1, 2 and 3).

(2) *Quotas-dependent implementation.* If no conditions are imposed on W, a direct computation of the generators of W will be less tractable. However, we show that with certain conditions on quotas system (b, c), we can circumvent this problem by first reducing the projection cone W in Lemma 1 into two separate cones. Then we can obtain an explicit characterization of the generators of each cone (Theorem 4).

3.1 Universal Implementation

The first main result (i.e., Theorem 1) of the paper will provide a sufficient condition on the projection cone W such that a complete description of the generators $\hat{W}$ can be found. The characterization depends on a class of integral matrices called *totally unimodular matrices.* This is a class of well-known and well-studied matrices (Schrijver, 1986). Formally, a matrix M is **totally unimodular** (TUM), if every nonsingular square submatrix has determinant equal to either -1 or $+1$. It is well known that there exists a polynomial time recognition algorithm for totally unimodular matrices; see Seymour (1980, 1981) and Schrijver (1986). In other words, one can easily verify whether a matrix is TUM or not.

Definition 6. *Let $(N, A, \mathcal{G}, T)$ be an implementation structure. We say $(N, A, \mathcal{G}, T)$ (or equivalently the projection cone W)* ***preserves total unimodularity****, if the* *constraint matrix $M(W)$ given by* (10) *is totally unimodular.*

Theorem 1 describes a key characterization of the implementability condition. It shows that if W preserves total unimodularity, then every generator (up to positive scaling) of W is a $\{0, \pm 1\}$ vector. That is,

$$(f, g, h) \in \hat{W} \implies (f, g, h) \in \{0, \pm 1\}^l \times \{0, +1\}^{2r}.$$

Hence every f corresponds to a sign function of some sets $S_i^+, S_i^- \subseteq T_i \times A$ for each $i \in N$. In this way we obtain a complete description of the set of implementable reduced forms.

We are ready to present the first major characterization result.

Theorem 1. (Universal Implementation) *Let $(N, A, \mathcal{G}, b, c, T, \lambda)$ be an implementation problem. Suppose the projection cone preserves total unimodularity. Then $Q \in \mathbb{R}^l$ is implementable if and only if Condition* (A) *holds.*

In general, Condition (A) is not separable in S^+ and S^-, since a priori both S^+ and S^- can take non-empty collections of sets, and sublinearity of β implies that condition for (S^+, S^-) is weakly tighter than a combination of separate conditions for S^+ and S^-.

Below we provide a sufficient condition on β that leads to a reduction of Condition (A) to separate expressions for ceiling constraints and floor constraints.

Corollary 1. *Suppose the projection cone W preserves total unimodularity and the generalized capacity function β is linear. Then $Q \in \mathbb{R}^l$ is implementable if and only if Condition* (B) *holds.*

A simple example of linearity is that there is no floor constraints except the usual non-negativity constraints (see Theorem 5 in Section 5).

Characterization of the projection cone. Now we will introduce our second and third major results (i.e., Theorems 2 and 3 below), which identify three classes of constraint structures $\mathcal{G}$ under which the projection cone preserves total unimodularity. If $\mathcal{G}$ is a hierarchy, then we show that the conic constraint matrix $M(W)$ consists of two laminars. Since the union of two laminars is total unimodular (Edmonds, 1970), we obtain that the projection cone preserves total unimodularity. However, when $\mathcal{G}$ is a bihierarchy, $M(W)$ has three laminars (i.e., C is two laminars and B is a laminar) to which the Edmonds theorem cannot be applied. Fortunately, we can still show that the projection cone preserves total unimodularity in this case if the set T_i for every $i \in N$ contains at most two elements, which is called binary.

Theorem 2. *Let $(N, A, \mathcal{G}, T)$ be an implementation structure. The projection cone preserves total unimodularity, if*

(1) $\mathcal{G}$ is a hierarchy.

(2) $\mathcal{G}$ is a bihierarchy and the set T_i for every $i \in N$ has at most two elements.

(3) $\mathcal{G}$ is a bihierarchy and the set T_i for every $i \in N$ has one element.

We remark that the case (3) of this theorem corresponds to Theorem 1 of Budish et al. (2013) on their universal implementation where they provide an alternative proof for this result. The proof is deferred to the Appendix.

Here we provide a sketch of proof for (2) of Theorem 2 for the bihierarchy and binary type case. Let C_i be the submatrix of C for the constraint sets in $\mathcal{G}_i$, $i = 1, 2$. Let B denote the constraint matrix of the reduced-form implementation equalities. Since total unimodularity is preserved by deleting unitary column[5] and duplicated column, and by transpose, we only need to show that the following matrix

$$M^* = \begin{bmatrix} C_1 \\ C_2 \\ B \end{bmatrix} \tag{12}$$

is totally unimodular.

We first prove the result for the problem with the standard constraints, i.e., for each hierarchy C_i, each column contains at most one 1 (in addition to the rows for singletons). Then we will show that all other cases can be reduced to the cases with the standard constraints. To show that M^* is TUM, we prove that it is sufficient to show that every square submatrix M' of M^* with exactly two nonzero entries per row and per column, the sum of the entries is a multiple of four, i.e., M' is balanced. We define a cycle P in M' as a series of changes of entries in M' with the two nonzero entries in each row and column adjacent. To complete the proof, we show that each cycle P in M' is even (i.e., the sum of entries in P is a multiple of four), given the bihierarchical and binary conditions.

[5]We call a column or row unitary if it has one nonzero entry.

	$(1a,t_1t_2)$	$(1b,t_1t_2)$	$(1b,t_1t_2')$	$(2b,t_1t_2')$	$(2b,t_1't_2')$	$(2c,t_1't_2')$	$(3c,t_1't_2')$	$(3c,t_1t_2')$	$(3a,t_1t_2')$	$(1a,t_1t_2')$
$(1,t_1t_2)$	1	1								
$(2,t_1't_2')$					1	1				
$(3,t_1t_2')$								1	1	
(a,t_1t_2')									1	1
(b,t_1t_2')			1	1						
$(c,t_1't_2')$						1	1			
$(1a,t_1)$	1									1
$(1b,t_1)$		1	1							
$(2b,t_2')$				1	1					
$(3c,t_3)$							1	1		

Table 1: Cycle P in submatrix M'.

$$P_{N\times A} = \begin{array}{c} \\ t_1t_2 \\ t_1t_2' \\ t_1't_2' \\ t_1't_2' \\ t_1t_2' \\ t_1t_2' \end{array} \begin{array}{c} \begin{array}{cccccc} 1a & 1b & 2b & 2c & 3c & 3a \end{array} \\ \begin{bmatrix} 1 & 1 & & & & \\ & 1 & 1 & & & \\ & & 1 & 1 & & \\ & & & 1 & 1 & \\ & & & & 1 & 1 \\ 1 & & & & & 1 \end{bmatrix} \end{array} \qquad P_T = \begin{array}{c} \\ 1b \\ 2b \\ 3c \\ 1a \end{array} \begin{array}{c} \begin{array}{cccc} t_1t_2 & t_1t_2' & t_1't_2' & t_1t_2' \end{array} \\ \begin{bmatrix} 1 & 1 & & \\ & 1 & 1 & \\ & & 1 & 1 \\ 1 & & & 1 \end{bmatrix} \end{array}$$

Table 2: Cycles on $N \times A$ and T.

For illustration, suppose $N = \{i_1, i_2, i_3\}$, $A = \{j_a, j_b, j_c\}$ and the constraints are canonical. Let $T_i = \{t_i, t_i'\}$, $i \in N$. Consider the following square submatrix M' of M^* with exactly two nonzero entries per row and per column in Table 1 (since M' has no index change for t_3, we abbreviate the coordinate t_3 for each row and column index).

First note that the non-zero entries in M' form a unique cycle P. For any pair of two 1s in each row, there are two types of possible changes of column indexes, either on $N \times A$ or on T. Then P induces two cycles $P_{N\times A}$ and P_T (see Table 2). The numbers of index changes in $P_{N\times A}$ and P_T are 6 and 4. So P is even and hence M' is balanced. In the Appendix, we show that this property holds for all such square submatrices.

Before presenting another major characterization result, we recall some concepts from graph theory; see Bondy and Murty (1976) and Schrijver (1986). A *cycle* in the graph $H = (N, E)$ is a sequence $(i_0, e_1, i_1, e_2, \cdots, i_{t-1}, e_t, i_t)$, where $i_0, i_1, \cdots, i_t$ are distinct vertices in N except $i_0 = i_t$ and $e_1, e_2, \cdots, e_t$ are distinct edges in E with $e_d = \{i_{d-1}, i_d\}$ for $d = 1, 2, \cdots, t$. A cycle is *odd* (even) if the number of edges on the cycle is odd (even). The next theorem shows that constraint structures with the adjacency property can also ensure the projection cone preserves total unimodularity.

Its proof is deferred to the Appendix.

Theorem 3. *Let $(N, A, \mathcal{G}, T)$ be an implementation structure. The projection cone preserves total unimodularity, if the graph $H = (N, E)$ does not contain any odd cycle and $\mathcal{G}$ is given by the set E of edges and T is binary.*

3.2 Quotas-dependent Implementation

To characterize the finite generators of the projection cone W requires some information on the implementation structure (i.e., total unimodularity). For characterization of implementability, however, it is not necessary to compute all generators of W, i.e., we only need to find the non-redundant generators with the tightest upper bounds in Lemma 1. Given paramodular property, we show that it is possible to characterize all $(f, g, h) \in W$ that attain the tightest upper bounds in Lemma 1 (not necessary the generators of W). This enables us to reduce the projection cone into two separate cones. It turns out each cone is totally unimodular. The sets of generators for the two cones then correspond to the implementability inequalities for the floor and ceiling constraints.[6]

Theorem 4. (Quotas-dependent Implementation) *Let $(N, A, \mathcal{G}, T, \lambda, b, c)$ be an implementation problem. Assume that (b, c) is paramodular. Then we have*

(1) The projection cone can be decomposed into two cones both preserving total unimodularity.

(2) $Q \in \mathbb{R}^l$ is implementable if and only if Condition (B) *holds.*

Theorem 4 generalizes in a meaningful way the classic characterization results of reduced form auctions (e.g., Border, 1991, 2007; Che et al., 2013) from a single good with multiple units to heterogeneous goods of which each good can have multiple units. It should be noted that Zheng (2021) is the first to obtain an elegant characterization similar to our Theorem 4. He presents his main result for a model with a continuum of types, while we present our Theorem 4 for a model with a finite number of types and our approach is totally different from his.[7] To obtain his characterization, he skillfully uses a separating hyperplane argument and constructs extreme

[6] We wish to thank anonymous referees for explicitly and implicitly raising the quotas-dependent implementation issue which led us to establish Theorem 4 and have related discussions.

[7] As pointed out by Zheng (2021), one can use a finite number of types to achieve results on a continuum of types by applying a technique as used by Che et al. (2013).

points of feasible reduced forms through a greedy algorithm. Zheng (2021) assumes that the goods can be divisible and hence allows real-valued quotas. It can be seen that our Theorem 4 continues to hold for real-valued quotas.

3.3 Necessity of conditions in Theorems 1 and 4

Che et al. (2013) have shown that paramodularity is necessary for their characterization, i.e., if the capacities fail paramodularity, then their characterization (and Theorem 4 in our paper) does not describe the set of the feasible reduced forms. We show that total unimodularity is necessary for our characterization Theorem 1. Here we revisit a very interesting example of Che et al. (2013) (their Remark 1).

Suppose there are three agents $N = \{i_1, i_2, i_3\}$, each independently and equally likely to be of type t_i^1 or t_i^2, and one type of good. The floor constraints are set zero and the ceiling constraints are given by

$$\begin{aligned} x_1 + x_2 + x_3 &\leq 6, \\ x_1 + x_2 &\leq 4, \\ x_1 + x_3 &\leq 4, \\ x_2 + x_3 &\leq 4, \\ x_1, x_2, x_3 &\leq 3. \end{aligned}$$

Observe that the problem is neither paramodular nor totally unimodular. However, all of the ceiling constraints are effective (i.e., $c(G) = \beta(G, \emptyset)$ for all $G \subseteq N$) and all extreme points are integral.[8]

Since there are only nonnegativity constraints in this example, the generalized capacity is linear and Condition (A) further reduces to Condition (B) (see also Theorem 5). Furthermore, since all of the constraints are effective, Condition (B) coincides with Condition (B′) in Che et al. (2013). Then we know from Che et al. (2013) that the characterization with (B′) is not valid: If we maximize a linear function $\sum_i Q_i(t_i^1) + 2Q_i(t_i^2)$ subject to Condition (B′) and compare the result to the maximum of the same objective function subject to the constraints (6) and (7), the value is strictly higher for maximization subject to (B′) ($18.375 > 18$). Hence, this example

[8] The system has 14 extreme points, including $(0, 0, 0)$, $(2, 2, 2)$, $(3, 1, 0)$, $(3, 0, 0)$ and the vectors by permuting coordinates in $(3, 1, 0)$, $(3, 0, 0)$.

also shows that total unimodularity is necessary for Theorem 1.

4 The Lift-and-Project Approach

In this section we present the mathematical method, i.e., the lift-and-project approach, to establish our characterization results introduced in the previous section. This is a powerful method in polyhedral combinatorics (Balas and Pulleyblank, 1983) and has been explored by Vohra (2013) in economics. The basic idea of the method goes as follows: The first step is to construct a linear system in some higher-dimensional space, or lifting. The second step is to obtain a linear system by properly projecting away previously added variables. Below we will keep the presentation as simple as possible while maintaining rigor.

First we describe a general lift-and-project method. Suppose we are given a polyhedron

$$Z = \{(x, y) \in \mathbb{R}^p \times \mathbb{R}^q \mid A^1x + B^1y = b^1,\ A^2x + B^2y \leq b^2\}, \tag{13}$$

where A^1, B^1, A^2, B^2 are matrices, b^1, b^2 are vectors. Let Y denote the projection of Z onto the subspace of y variables, that is,

$$Y = \{y \in \mathbb{R}^q \mid \text{ there exists } x \in \mathbb{R}^p \text{ such that } (x, y) \in Z\}. \tag{14}$$

We wish to obtain a linear system whose solution set is Y.

We define the projection cone of Z by

$$P = \{(f, g) \in \mathbb{R}^r \times \mathbb{R}^s \mid f^\top A^1 + g^\top A^2 = 0, g \geq 0\}. \tag{15}$$

Let $\hat{P}$ be any finite set of generators of P. The following lemma shows that finding the linear inequalities that define Y reduces to finding finite generators of P.

Lemma 2. *The projection of the polyhedron Z given by* (13) *onto y is given by*

$$Y = \{y \in \mathbb{R}^q \mid (f^\top B^1 + g^\top B^2)y \leq f^\top b^1 + g^\top b^2 \textit{ for all } (f, g) \in \hat{P}\}. \tag{16}$$

Characterizing the generators of P appears to be difficult in general. The following lemma provides a sufficient condition on P such that a complete description of the generators can be found. The lemma is developed from and slightly more general than a result in Hoffman (1976), where the result there is proved for pointed cones. We show that the same result holds also for non-pointed cones.

Lemma 3. *If the constraint matrix of the projection cone P given by* (15) *is totally unimodular, then P is generated by a set $\hat{P}$ of $\{0, \pm 1\}$ generators.*

Remark 1. If P is pointed, the set of extreme rays provides a unique (up to positive scaling) minimal set of generators. While Lemma 2 shows that the extreme rays of P provide a complete description of Y, certain extreme rays may be redundant.

Remark 2. Lemma 3 is useful on its own since it identifies finitely many linear inequalities. In particular, it implies that we can pick $\hat{P}$ to be all $\{0, \pm 1\}$ vectors in P, i.e., $P \cap \{0, \pm 1\}^{r+s}$.

Remark 3. Hoffman (1976) also showed a converse to Lemma 3, i.e., if the constraint matrix of P is not totally unimodular, then there exists some extreme ray that is not $\{0, \pm 1\}$.

To prove Theorem 1, the first step is to apply the lift-and-project method to our implementation problem. First note that if the projection cone preserves TUM, then the constraint matrix of $\mathcal{G}$ is also TUM, as every submatrix of a TUM matrix is TUM. Hence, the set of feasible ex post random allocations has a linear characterization defined by Condition (1). Then the linear constraints F in (8) describe the implementation system and hence Lemma 1 applies to F. Since the projection cone W is totally unimodular, applying Lemma 3 to W we obtain Lemma 4 below, which characterizes the generators of the projection cone in Theorem 1.

Lemma 4. *Suppose W preserves total unimodularity. Then $y \in \mathbb{R}^l$ is implementable if and only if*

$$f^\top y \leq g^\top c^\lambda - h^\top b^\lambda \quad \textit{for all } (f, g, h) \in \hat{W}, \tag{17}$$

where $\hat{W}$ consists of generators of W given by (9) *with entries $-1, 0,$ and $+1$.*

To prove Theorem 4, first note that when (b, c) is paramodularity, integral polymatroids imply that the linear constraints (6) completely describe the set of feasible random allocations. Then the linear constraints F in (8) describe the implementation system and hence Lemma 1 applies to F. We next characterize all $(f, g, h) \in W$ that attain the tightest upper bounds in Lemma 1, not necessarily generators of W. This leads to a reduction of the projection cone into two separate cones, which can be shown to be totally unimodular. We then characterize the generators of each cone.

Below we provide some intuition for this reduction at one type profile. For any given $f \in \mathbb{R}^{N\times A}$, we maximize $f^\top x$ subject to constraints (6). Consider its dual problem

$$\min g^\top c - h^\top b \ \text{ s.t. } \ g, h \geq 0, (g^\top - h^\top)\chi = f, \tag{18}$$

where χ denotes the matrix whose rows are the characteristic vectors χ^U of the subsets $U \subseteq N \times A$.

For any optimal solution (g, h) to (18), let $\mathtt{supp}(g)$ and $\mathtt{supp}(h)$ denote the support of the solution. We say the pair (b, c) is totally dual laminar (TDL) if for every f with finite optimum, some optimal dual solution to (18) is laminar, i.e., $\mathtt{supp}(g) \cup \mathtt{supp}(h)$ is a laminar. It is well known that paramodularity implies TDL. Moreover, Theorem 6 in Frank et al. (2013) shows that we can gradually make an optimal dual solution more and more structured.

Lemma 5. *Suppose the pair (b, c) is TDL. Then the problem (18) has an optimum (g, h) such that $\mathtt{supp}(g)$ and $\mathtt{supp}(h)$ are laminar families on disjoint ground sets. Furthermore, $\mathtt{supp}(g)$ and $\mathtt{supp}(h)$ are chain families on disjoint ground sets.*

Intuitively, Lemma 5 suggests that floor and ceiling constraints can be treated separately if g and h have disjoint ground sets. Below we extend this pointwise analysis to the entire implementation system. We need to introduce a definition. Given a nonempty ground set K, let F be a family of subsets of the set K. We say that F is a chain on K if we have either $G \subseteq G'$ or $G' \subseteq G$ for any $G, G' \in F$. Now let $\mathcal{C} = (\mathcal{C}^t)_{t\in T}$ be a profile of chains, where every $\mathcal{C}^t$ is a chain on $N \times A$. Let $\mathbf{C}^T$ denote all possible profiles of chains. For every pair of profiles of chains $(\mathcal{C}_1, \mathcal{C}_2)$, define the cone

$$W(\mathcal{C}_1, \mathcal{C}_2) = \{(f, g, h) \in W |\ g, h \text{ have supports } \mathcal{C}_1 \text{ and } \mathcal{C}_2\}.$$

From Lemma 5, if (f, g, h) obtains a tightest bound in Lemma 1, then $(f, g, h) \in W(\mathcal{C}_1, \mathcal{C}_2)$ for some pair of profiles of chains $(\mathcal{C}_1, \mathcal{C}_2)$ where $\mathcal{C}_1$ and $\mathcal{C}_2$ have disjoint ground sets (i.e., each $\mathcal{C}_1^t$ and $\mathcal{C}_2^t$ have disjoint ground sets). This further implies that a reduction of $W(\mathcal{C}_1, \mathcal{C}_2)$ into two separate cones with ceiling and floor constraints is possible. To state the following lemma, we define two separate cones

$$W^+(\mathcal{C}_1) = W(\mathcal{C}_1, \emptyset) \cap \{f \geq 0, h = 0\},$$
$$W^-(\mathcal{C}_2) = W(\emptyset, \mathcal{C}_2) \cap \{f \leq 0, g = 0\}.$$

Lemma 6. *Let $(N, A, \mathcal{G}, T, \lambda, b, c)$ be an implementation problem. Suppose quotas are paramodular. Then $y \in \mathbb{R}^l$ is implementable if and only if*

$$f^\top y \leq g^\top c \text{ for all } (f, g) \in \cup_{\mathcal{C}_1 \in \mathbf{C}^T} W^+(\mathcal{C}_1), \tag{19}$$

and

$$-f^\top y \geq h^\top b \text{ for all } (f, h) \in \cup_{\mathcal{C}_2 \in \mathbf{C}^T} W^+(\mathcal{C}_2). \tag{20}$$

From Lemma 6, each $\mathcal{C}_1$ is a laminar on $N \times A \times T$ since each $\mathcal{C}_1^t$ is a profile of chains on $N \times A$. Also, the constraint sets in reduced-form submatrix B is a laminar on $N \times A \times T$. By Edmonds (1970), the union of two laminars forms a totally unimodular matrix. Then Lemma 3 implies that the cone $W^+(\mathcal{C}_1)$ is generated by extreme rays with $0, 1$ entries. Similarly, $W^-(\mathcal{C}_2)$ is generated by extreme rays with $0, -1$ entries. The next lemma provides a characterization of the sets of extreme rays of $W^+(\mathcal{C}_1)$ and $W^-(\mathcal{C}_2)$. Define $\Gamma(S) = \{t \in T \mid (t_i, j) \in S_i, \text{ for some } i \in N\}$.

Lemma 7. (1) *Let $\mathcal{C}_1$ be any profile of chains. (f, g) is an extreme ray of the set $W^+(\mathcal{C}_1)$ if and only if*

$$f(i, t_i, j) = \begin{cases} +1, & \text{if } (t_i, j) \in S_i, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$g(U, t) = \begin{cases} +1, & \text{if } t \in \Gamma(S) \text{ and } U = I(t, S), \\ 0, & \text{otherwise.} \end{cases}$$

(2) *Let $\mathcal{C}_2$ be any profile of chains. (f, h) is an extreme ray of the set $W^-(\mathcal{C}_2)$ if and only if*

$$f(i, t_i, j) = \begin{cases} -1, & \text{if } (t_i, j) \in S_i, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$h(U, t) = \begin{cases} +1, & \text{if } t \in \Gamma(S) \text{ and } U = I(t, S), \\ 0, & \text{otherwise.} \end{cases}$$

5 Applications

In this section, we provide several applications of practical interest. The examples include systems of multiple suppliers, assignment markets, allocation of club goods, assignment of radio spectrum licenses, bilateral trade, and compromise problem.

5.1 Systems of multiple suppliers

Consider the automobile manufacturing industry with multiple suppliers. Each supplier (e.g., Toyota) decides how to set its capacities for different models (i.e., Corolla, Camry, etc) and sell its products. While the maximum capacity for each model is fixed in the short run, the supplier can adjust its maximum capacity in the long run, in which case all possible capacities may arise.

Let L be a finite set of suppliers or car makers. Each supplier $l \in L$ can produce a finite set A^l of models. Note that $A^l \cap A^{l'} = \emptyset$ for any $l \neq l'$. Denote $A = \cup_{l \in L} A^l$. For each model $j \in A$, the cost to install capacity $k_j \geq 0$ is $C_j(k_j)$. In society there are two types of consumers (e.g., families), type X and type Y, with $X \cup Y = N$. A type X family wishes to buy at most one car, while a type Y family wants to buy at most two cars of different models.

Given a capacity profile $k \in \mathbb{R}_+^A$, the resource constraints for the suppliers are given by

$$0 \leq x(N, j) \leq k_j \text{ for all } j \in A. \tag{21}$$

The constraints for families of two types are given by

$$0 \leq x(i, A) \leq 1 \text{ for all } i \in X, \tag{22}$$

$$0 \leq x(i, A) \leq 2 \text{ for all } i \in Y, \tag{23}$$

$$x(i, j) \in \{0, 1\} \text{ for all } i \in N, j \in A. \tag{24}$$

The type of each family is public information. We have the following result.

Lemma 8. *Conditions* (21)-(24) *are bihierarchical.*

Note that each supplier l's decisions are usually made at two stages. At the first stage the supplier chooses the capacities $k^l \in \mathbb{R}_+^{A^l}$ for all types of its products. The capacities determine a set of resource feasible allocations satisfying (21). At the second stage, the supplier chooses feasible allocations to assign its products. Let $R^l(p; k^l)$ denote the supplier's expected payoff (i.e., revenue) from a feasible allocation rule p and let $R^l(k^l)$ denote the continuation payoff from any revenue-maximizing allocation rule. The supplier's problem at the first stage is to choose capacity profile k^l to maximize the profit. In this case, all possible capacities can arise in the second stage and a characterization of feasibility with arbitrary quotas is desirable.

Similar problems can arise from other manufacturing industries. Another interesting application concerns systems of policy-determined quotas. Markets can be regulated by distributional policies in many situations. A typical example is the regional cap in Japanese residency matching program (e.g., Kamada and Kojima, 2015). In school choice problems, the regional cap applies to each district (e.g., Biró et al., 2010, Ehlers et al., 2014, Fragiadakis and Troyan, 2016).

For the revenue maximization design with transfers, quasi-linear preferences allow us to separate out transfer and allocation rule, by a revenue equivalence argument. The literature has been devoted to identifying sufficient conditions on the type space for all allocation rules from a certain class to satisfy revenue equivalence. It can be shown that our model satisfies the sufficient conditions identified in the literature (Heydenreich et al. (2009), Theorem 8). We refer to Heydenreich et al. (2009) for an in-depth study.

5.2 Assignment markets

We now discuss the implementation problem for the classical model of assignment markets studied by Koopmans and Beckmann (1957), Shapley and Shubik (1971), Crawford and Knoer (1981), Demange et al. (1986), and Bogomolnaia and Moulin (2001). For the classical model it is typically assumed that every buyer demands one item and every seller sells one item. Here we drop this assumption. In the following we examine two important cases: Ordinal preferences and cardinal preferences.

With the constraints $x \geq 0$, the generalized capacity function β is linear, since $\beta(\emptyset, V) = 0$ for all V and $\beta(U, V) = \beta(U, \emptyset)$ for all $(U, V) \in 3^{N \times A}$. Hence Corollary 1 implies that the floor and ceiling constraints are separable and we obtain the following result.

Theorem 5. *Let $b = (0, \cdots, 0)$ and let c be any vector in $\mathbb{Z}_+^{|\mathcal{G}|}$. Suppose the projection cone preserves total unimodularity. Then Q is implementable if and only if for all i*

and $(t_i, j) \in T_i \times A$, $Q_i(t_i, j) \geq 0$, *and for all* $S_i \subseteq T_i \times A$, $i \in N$,

$$\sum_{i \in N} \sum_{(t_i, j) \in S_i} Q_i(t_i, j)\lambda_i(t_i) \leq \sum_{t \in \Gamma(S)} \lambda(t)\beta((I(t, S)), \emptyset). \tag{25}$$

In the theorem, if we set the supply vector $c = (1, \cdots, 1)$, the model reduces to the classical assignment market. It is worth mentioning that computing the generalized capacity function $\bar{\beta}(U) := \beta(U, \emptyset)$ for each $U \subseteq N \times A$ is a maximum weight matching problem. We provide three useful representations of $\bar{\beta}(U)$. From the definition of $\bar{\beta}$, we first obtain the support function representation:

$$\bar{\beta}(U) = \max\{\chi^U \cdot x : x \geq 0, \sum_{j \in A} x_{ij} \leq 1, \sum_{i \in N} x_{ij} \leq 1\}$$

where $\chi^U_{ij} \in \{0, 1\}$ can be interpreted as the surplus for a buyer-seller pair (i, j) and $\bar{\beta}$ is the efficient social surplus. It is well known that the solution to this problem can be computed efficiently through the Hungarian method (Demange et al., 1986). By strong duality of linear programming (e.g., Shapley and Shubik, 1971),

$$\bar{\beta}(U) = \min\{\sum_{i \in N} s_i + \sum_{j \in A} p_j : s_i, p_j \geq 0, s_i + p_j \geq \chi^U_{ij}\}$$

where s_i and p_j are the payoffs for buyer i and seller j. From the the König-Egeváry theorem (see Schrijver, 1986), we can further obtain the following minimum cover representation

$$\bar{\beta}(U) = \min\{|\mathcal{P}| : \mathcal{P} \subseteq \mathcal{G} \text{ is a cover of } U\}.$$

Intuitively, let $M^U \in \mathbb{R}^{N \times A}$ denote the matrix representation of U. Then the König-Egeváry theorem asserts that the maximal number of nonzero elements of M^U with no two of these elements on a line is equal to the minimal number of lines of M^U that contain all the nonzero elements of M^U. Without loss of generality, let us assume that $m = n$. Let $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$ be a permutation and P_σ be a permutation matrix of order n where $p_{1\sigma(1)} = p_{2\sigma(2)} = \cdots = p_{n\sigma(n)} = 1$. Denote

$$P_\sigma \circ M^U = M^U_{1\sigma(1)} + \cdots + M^U_{n\sigma(n)}$$

the diagonal sum of M^U corresponding to P_σ. We obtain the following maximum-sum diagonal representation

$$\bar{\beta}(U) = \max\{P_\sigma \circ M^U : P_\sigma \text{ is a permutation matrix of order } n\}.$$

Example 6. *Suppose* $N = \{i_1, i_2, i_3\}$, $A = \{j_1, j_2, j_3\}$ *and* $U = \{(i_1, j_1), (i_1, j_3), (i_2, j_2)\}$. *Then* $\bar{\beta}(U) = 2$. *The minimum covers are either* $\{$*row* i_1, *row* $i_2\}$ *or* $\{$*row* i_1, *column* $j_2\}$. *The maximum diagonal sum is equal to* $M^U_{11} + M^U_{22} = M^U_{13} + M^U_{22} = 2$.

The case of ordinal types. We first consider the model with ordinal preferences. For each agent i, a type $t_i \in T_i$ denotes a strict preference over A, denoted by $\succ_{t_i}$. For any good j and type t_i, let $B(a, t_i) := \{j\} \cup \{k \in A : k \succ_{t_i} j\}$. For any random outcomes x and x', we say x first-order-stochastically-dominates (FOSD) x' according to t_i ($x \succeq_{t_i} x'$) if $\sum_{j \in B(k,t_i)} x(i,j) \geq \sum_{j \in B(k,t_i)} x'(i,j)$, for all $k = 1, ..., m$. An interim allocation rule Q is ordinally Bayesian incentive compatible (OBIC) if for all $i \in N$ we have

$$Q_i(t_i) \succeq_{t_i} Q_i(t_i') \text{ for all } t_i, t_i'. \tag{26}$$

The following result provides a characterization of the set of ordinally Bayesian incentive compatible allocations for random assignment problems.

Corollary 2. *Let* T_i, $i \in N$ *be type sets of ordinal preferences. The set of OBIC and implementable allocation rules is given by Conditions* (25) *and* (26).

Notice that when there are two types of objects, i.e., $|A| = 2$, each agent has a binary type set (i.e., $T_i = A$). From Theorem 2 the projection cone preserves total unimodularity.

The case of cardinal types. Now we turn to the model with cardinal utilities. Suppose each agent i has a set T_i of types that represent her payoff vectors over the objects, i.e., $t_i \in \mathbb{R}^A$. Assume the payoffs over objects are additive. For any random outcomes x and x', we say agent i with type t_i prefers x to x' if $x(i) \cdot t_i \geq x'(i) \cdot t_i$. An interim allocation rule Q is incentive compatible if for all $i \in N$,

$$Q_i(t_i) \cdot t_i \geq Q_i(t_i') \cdot t_i \text{ for all } t_i, t_i'. \tag{27}$$

Corollary 3. *Let* T_i, $i \in N$ *be type sets of cardinal utilities. The set of incentive compatible and implementable allocation rules is given by Conditions* (25) *and* (27).

5.3 Allocation of club goods

Consider an allocation problem of excludable public goods (club goods) where any agent's consumption can be restricted at zero cost to any level below the total quantity produced. Examples range from airports, zoos, museums to TV channels, database, and different community clubs. We show how the conic approach can be applied to this problem, which allows a joint implementation of public and private goods.

Suppose that a monopolist can produce $y \in [0, 1]$ units of the excludable public good by incurring costs $cy \geq 0$. There are two agents 1 and 2. Each agent i has a consumption level of the public good $q_i \in [0, y]$, with the strict inequality holding whenever exclusion is exercised. Since the monopolist derives no intrinsic utility from the public good, we can set $y = \max\{q_1, q_2\}$. Suppose each agent has $l > 0$ units of the private good and a payment $p_i \in [0, l]$. We say an allocation (q, p) is feasible if the project's input covers the cost

$$p_1 + p_2 \geq c \max\{q_1, q_2\}. \tag{28}$$

We can rewrite this inequality into two separate inequalities:

$$p_1 + p_2 - cq_i \geq 0, \text{ for each } i. \tag{29}$$

Denote $x_i^a = cq_i$ and $x_i^b = l - p_i$ as the allocation of the public good and the residual private good for agent i. The problem fits our model with two types of goods a and b. Observe that the constraint structure of (29) is non-canonical but bihierarchical.

5.4 Allocation of radio spectrum licenses

Suppose that some radio spectrum licenses are allocated to seven service providers $N = \{i_1, i_2, ..., i_7\}$ located in different geographic regions. Let $H = (N, E)$ denote the interference graph for the providers, where a pair $\{i, j\}$ of providers form an edge if and only if i and j are close to each other and can have radio frequency interference when i and j are both on-air. An example of the interference graph H is given by Figure 1.

The licenses are identical but labelled differently and have multiple units. For each pair of neighbors $\{i, j\} \in E$, at most one provider can be allocated a license

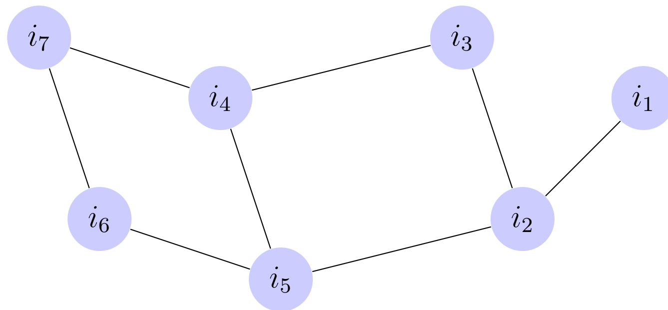

Figure 1: Radio frequency interference graph H

to avoid radio interference. The following system of linear inequalities describes all possible outcomes:

$$\begin{aligned} 0 \le x(i) + x(j) \le 1 \quad & \text{for every} \quad \{i, j\} \in E, \\ x(i), x(j) \in \{0, 1\} \quad & \text{for every} \quad \{i, j\} \in E. \end{aligned}$$

Here we give a simple example for the purpose of illustration. We refer to Rothkopf et al. (1998) and Milgrom and Segal (2020) for detailed studies on the topic.

5.5 Bilateral trade

We revisit the classic bilateral trade problem of Myerson and Satterthwaite (1983). The probability simplex constraint assumes that for any type profile of the buyer i_1 and the seller i_2, the probabilities of each agent obtaining the object satisfy $p^1 + p^2 = 1$ and $p \geq 0$. In many scenarios, the agents may be forced to take the status quo if their type profiles are incompatible. In other practical settings, the players can deliberately choose to implement the status quo for some profiles (e.g., ex post efficiency).

To incorporate such side constraints on allocation rules, we introduce an extended problem with side constraints. Let $(T_1 \cup T_2; E)$ be a bipartite graph on the type sets T_1 and T_2, where $(t_1, t_2) \in E$ means that there is an edge between $t_1 \in T_1$ and $t_2 \in T_2$, i.e., (t_1, t_2) is a compatible pair ex post. We say a feasible allocation rule p is E-feasible if $p^2(t) = 1$ for all $t \notin E$. The problem with no side constraints corresponds to a complete graph on $T_1 \cup T_2$.

We show how the conic method can be applied to this problem. Observe that the problem corresponds to the case of our model where $|A| = 1$, $\mathcal{G} = \{N\}$, $b(N) =$

$c(N) = 1$, and $b(\{i\}) = 0$. Moreover in B and C only the entries restricted to $t \in E$ are non-zeros. The equal ceiling and floor constraint implies that g is free in the set W given by (9). By eliminating $h \geq 0$, the projection cone reduces to

$$W = \{(f, g) \in \mathbb{R}^{l+|E|} \mid -f^\top B + g^\top C \geq 0\}.$$

We characterize the generators of the projection cone for this bilateral trade problem. In contrast to Theorem 5 whose projection cone is *pointed*, we show that the projection cone of a bilateral trade is *non-pointed.*

Lemma 9. *For bilateral trade with constraints* $(T_1 \cup T_2; E)$*, we have: (1)* W *is non-pointed. (2)* $M(W)$ *is TUM.*

The next result gives a detailed characterization of the projection cone concerning bilateral trade.

Lemma 10. *For bilateral trade with constraints* $(T_1 \cup T_2; E)$*, if* (f, g) *is a* $\{0, \pm 1\}$ *generator of* W*, then one of the following conditions holds:*

(1) $(f, g) = (1, \ldots, 1)$ *or* $(f, g) = (-1, \ldots, -1)$.

(2) $f(i, t_i) = -1$ *for a unique* (i, t_i) *and 0 otherwise, and* $g = 0$.

(3) (f, g) *is the incidence vector of some* $(S, \Gamma(S))$*, where* $S_i \subseteq T_i$ *for all* $i \in N$ *and* $\Gamma(S) = \{t \in E \mid t_i \in S_i \text{ for some } i \in N\}$*. That is,*

$$f(i, t_i) = \begin{cases} +1, & \text{if } t_i \in S_i, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$g(t) = \begin{cases} +1, & \text{if } t \in \Gamma(S), \\ 0, & \text{otherwise.} \end{cases}$$

The generators in condition (1) of this lemma correspond to a basis of the linearity space of the projection cone defined by $\{x \in W : Mx = 0\}$. These generators correspond to the equality constraint in the implementability condition:

$$\sum_{i=1,2} \sum_{t_i \in T_i} Q_i^i(t_i) \lambda_i(t_i) = \sum_{t \in E} \lambda(t). \tag{30}$$

Note that if the projection cone is non-pointed, the linear description is not unique and (30) can be used to generate different descriptions. For bilateral trade, as both the seller and the buyer are interested in the interim expected probability of trade (i.e., $Q_i^1(t_i)$), we can obtain the following characterization result.

Proposition 3. *For bilateral trade with constraints* $(T_1 \cup T_2; E)$, $Q \in \mathbb{R}_+^l$ *given by* (30) *is implementable if and only if*

$$\sum_{t_1 \in T_1} Q_1^1(t_1)\lambda_1(t_1) - \sum_{t_2 \in T_2} Q_2^1(t_2)\lambda_2(t_2) = 0, \tag{31}$$

and for all $S_1 \subseteq T_1$, $S_2 \subseteq T_2$,

$$\sum_{t_1 \in S_1} Q_1^1(t_1)\lambda_1(t_1) - \sum_{t_2 \in S_2} Q_2^1(t_2)\lambda_2(t_2) \leq \lambda((S_1 \times S_2^c) \cap E). \tag{32}$$

That is, at the ex ante stage, the two players must have the same expectations on the probability of trade, and the difference in the seller's and the buyer's interim probabilities of trade for any set of types $S_1 \times S_2$ *cannot be too distinct.*

We provide some interpretation of the implementability conditions, by comparing an auction with a bilateral trade. In Border's theorem, only coefficients "+1" appear in the linear inequalities. In contrast, the implementability condition for bilateral trade here contains coefficients not only "+1" but also "−1" in the linear inequalities. The interpretation for this result is intuitive: In Myerson (1981), the "+1" coefficient means that if for buyer 1 the expected probability of winning becomes higher, then for buyer 2 the expected probability of winning must be lower as the buyers are competing for the probabilities of winning. In Myerson and Satterthwaite (1983), however, the "+1" coefficient refers to that for each player her expected probability of trade for some types is obtained by summing up her interim probabilities of trade for these types, the "−1" coefficient means that the difference between the seller's ex ante expected probability of trade and the buyer's cannot be too large. This is because increasing the probability that trade occurs would increase both players' expected probabilities of trade.

5.6 Compromise

Our final application concerns the compromise problem studied by Börgers and Postl (2009) which is totally different from the previous ones. In their model there are two players (i.e., $N = \{1, 2\}$) and three alternatives (i.e., $A = \{a_0, a_1, a_2\}$).[9] The players have opposite preferences: $a_1 \succ_1 a_0 \succ_1 a_2$ and $a_2 \succ_2 a_0 \succ_2 a_1$. That is, the best alternative to one player is the worst alternative to the other player. We normalize $u_i(a_0) = 0$ for $i = 1, 2$. Each player i has private information about her payoffs on a_1 and a_2, given by a type $t_i \in T_i$.[10]

While the number of types in Börgers and Postl (2009) can be any positive integer, for illustrative purpose we focus on the case of two types, i.e., $|T_1| = |T_2| = 2$. In this problem a feasible allocation rule $q : T \to \Delta(A)$ assigns each type profile a lottery over alternatives. Hence for each player with type t_i, the reduced form allocation probability is multidimensional: $Q_i(t_i) = (Q_i(t_i, a_1), Q_i(t_i, a_2))$.

To apply our approach in this setting, let B be an incidence matrix where each row is indexed by (i, a, t_i), each column is indexed by (a, t), and the entry in row (i, a, t_i) and column (a', t') is 1 if $a = a'$ and $t_i = t'_i$ and 0 otherwise. Let C be an incidence matrix where each row is indexed by t, each column is indexed by (a, t), and the entry in row t and column (a', t') is 1 if $t = t'$ and 0 otherwise. The projection cone is given by

$$W = \{(f, g) \mid -f^\top B + g^\top C \geq 0\}.$$

We will prove that when each player has two types, the projection cone is non-pointed and the constraint matrix of the projection cone is totally unimodular.

Lemma 11. *For compromise with binary type sets, we have: (1) W is non-pointed. (2) $M(W)$ is TUM.*

It is worth noting that while $M(W)$ does not form two laminars (it contains three laminars), which differs from the bilateral trade problem, the constraint matrix remains totally unimodular.

The above result implies that the projection cone is generated by $\{0, \pm 1\}$ vectors. To present the implementability condition, let $K = \{a_1, a_2\}$. It can be verified that

[9] See Lang and Mishra (2023) for a recent study on a related voting problem with two public alternatives.

[10] Note that Börgers and Postl (2009) normalize $u_i(a_i) = 1$ and $u_i(a_j) = 0$ and assume each player has private information about her payoff on the compromise alternative a_0, i.e., $u_i(a_0) = t_i$. The reduced-form implementation problem is the same irrespective of the normalizations.

the implementability inequalities are given by: for all $a \in K$,

$$\sum_{t_1 \in T_1} Q_1(t_1, a)\lambda_1(t_1) - \sum_{t_2 \in T_2} Q_2(t_2, a)\lambda_2(t_2) = 0, \tag{33}$$

and for all $G \subseteq K$, $S_1 \subseteq T_1$, $S_2 \subseteq T_2$,

$$\sum_{a \in G} \sum_{t_1 \in S_1} Q_1(t_1, a)\lambda_1(t_1) - \sum_{t_2 \in S_2} Q_2(t_2, a)\lambda_2(t_2) \leq \lambda(S_1 \times S_2^c). \tag{34}$$

6 Concluding Remarks

Many practical problems and markets face various complex distributional constraints going beyond the traditional ceiling or capacity constraints. In this paper, we have studied the implementation problem of reduced-form allocation of multiple heterogeneous indivisible objects to many agents with distributional constraints. In our model, objects can be private goods or public/club goods, every agent may demand several objects and has private information over her preferences. Her private information is described by a finite set of types. Distributional constraints are described by a variety of families of pairs of agent and object.

Using a lift-and-project method, we have been able to obtain a conic approach for studying two major classes of implementability problems: Universal implementation and quotas-dependent implementation. We have demonstrated how the approach allows for a unified treatment of different classes of problems. We have succeeded in identifying a fundamental and very general condition called total unimodularity and establishing four general characterization results on implementation. Total unimodularity reflects the essential property of the class of well-studied totally unimodular matrices. Analyzing these matrices offers also interesting criteria that can be explored to classify different classes of economic problems. For each problem, the main task is to check whether its projection cone preserves total unimodularity or not. In fact, we have proved that four large classes of constraint structures: hierarchies, bihierarchies, adjacency, and paramodularity each can ensure total unimodularity. We have also provided several applications of practical interest including systems of multiple suppliers, assignment markets, allocation of club goods, assignment of radio spectrum licenses, bilateral trade, and compromise problem.

We hope this study has shed new light on implementation of reduced-form allocation and will provide a useful and necessary basis for further study on many complex real life resource allocation problems.

A Appendix

A.1 Illustration for Condition (A)

Example 7. (Example 2 continued) *Suppose that there are two students* $N = \{i_1, i_2\}$, *one compulsory course* c_1 *and three optional courses* $O = \{o_1, o_2, o_3\}$, *and* $T_1 = \{t_1^1, t_1^2\}$ *and* $T_2 = \{t_2^1, t_2^2\}$. *Every student is required to take* c_1, *and at least one and at most two of optional courses. Each optional course is required to enroll at least one and at most two of students. Notice that the compulsory course can be treated separately and we restrict our attention to optional courses. Pick* $S_1^+ = \{(t_1^2, o_2)\}$, $S_2^+ = \{(t_2^1, o_1), (t_2^1, o_2), (t_2^1, o_3)\}$, $S_1^- = \{(t_1^1, o_1), (t_1^1, o_2), (t_1^1, o_3)\}$, $S_2^- = \{(t_2^2, o_2)\}$. *We calculate* β *in Table 3. In particular, for* (t_1^1, t_2^1) *consider the following problem:*

Table 3: Parameters in Example 7.

t	$I(t, S^+)$	$I(t, S^-)$	β
(t_1^1, t_2^1)	$\{(i_2, o_1), (i_2, o_2), (i_2, o_3)\}$	$\{(i_1, o_1), (i_1, o_2), (i_1, o_3)\}$	$c(\{i_2\} \times O) - b(\{i_1\} \times O)$
(t_1^1, t_2^2)	$\{\emptyset\}$	$\{(i_1, o_1), (i_1, o_2), (i_1, o_3), (i_2, o_2)\}$	$-b(\{i_1\} \times O)$
(t_1^2, t_2^1)	$\{(i_1, o_2), (i_2, o_1), (i_2, o_2), (i_2, o_3)\}$	$\{\emptyset\}$	$c(\{i_1\} \times \{o_2\}) + c(\{i_2\} \times O)$
(t_1^2, t_2^2)	$\{(i_1, o_2)\}$	$\{(i_2, o_2)\}$	$c(\{i_1\} \times \{o_2\})$

$$\begin{aligned} \max\ & x(i_2, o_1) + x(i_2, o_2) + x(i_2, o_3) - x(i_1, o_1) - x(i_1, o_2) - x(i_1, o_3) \\ s.t.\ & 1 \le x(i, O) \le 2,\ i \in N, \\ & 1 \le x(N, o) \le 2,\ j \in O,\ \ 0 \le x \le 1. \end{aligned}$$

An optimal solution is given by $x^*(i_1, o_1) = x^*(i_2, o_1) = x^*(i_2, o_3) = 1$, *and* $x^*(i, o) = 0$ *otherwise. Note that at* x^* *the floor constraint for* i_1 *and the ceiling constraint for* i_2, *and the floor constraints for all optional courses are all binding. Similarly, we calculate* β *for the other type profiles. The corresponding implementability condition is given by*

$$Q_{1,2}^2 + Q_{2,1}^1 + Q_{2,1}^2 + Q_{2,1}^3 - Q_{1,1}^1 - Q_{1,1}^2 - Q_{1,1}^3 - Q_{2,2}^2 \le \lambda_{11} - \lambda_{12} + 3\lambda_{21} + \lambda_{22},$$

where $Q^j_{i,k} := Q_i(t^k_i, o_j)\lambda_i(t^k_i)$ *and* $\lambda_{kl} := \lambda(t^k_1, t^l_2)$.

A.2 Proof of Theorem 1

Proof of Theorem 1. By Lemma 4, if (f, g, h) is a generator of W, then f corresponds to a sign function of some sets $S^+_i, S^-_i \subseteq T_i \times A$ for each $i \in N$. The projection of f onto $t \in T$, denoted by f_t, is given by

$$sign^{I(t,S^+),I(t,S^-)}(i,j) = \begin{cases} +1 & \text{if } (i,j) \in I(t,S^+), \\ -1 & \text{if } (i,j) \in I(t,S^-), \\ 0 & \text{otherwise.} \end{cases}$$

To obtain the tightest upper bound in Lemma 4 for each such f, we minimize $g^\top c^\lambda - h^\top b^\lambda$ subject to $(f, g, h) \in W$. For every $t \in T$, the pointwise optimization is given by

$$\min g_t^\top c - h_t^\top b \;\text{ s.t. }\; g_t, h_t \geq 0, (g_t^\top - h_t^\top)\chi = sign^{I(t,S^+),I(t,S^-)},$$

where g_t, h_t are the projection of g, h onto t and χ denotes the matrix whose rows are the characteristic vectors χ^U of the subsets $U \in \mathcal{G}$. By a strong duality of linear programming, we have the value of this minimization problem is equal to $\beta(I(t,S^+), I(t,S^-))$. □

A.3 Proof of Theorem 2

Proof of Theorem 2. (1) If $\mathcal{G}$ is one hierarchy, then C is a laminar on $N \times A \times T$, since C restricted to each $t \in T$ is a laminar. Notice that B is also a laminar on $N \times A \times T$. From Edmonds (1970), the union of two laminars is TUM.

(2) Without loss of any generality, we consider a case where each agent has exactly two types. Let M^* be given in condition (12). Since TUM is preserved by deleting unitary rows, it is without loss to consider non-singleton constraints in M^*. We first show that result holds for standard bihierarchies where each column contains two ones.

(a) Standard constraints. In this case, each column of M^* contains exactly three 1s. Note that if a matrix A with at most three nonzero entries in each column, then A is TUM if and only if each submatrix of A with at most two nonzero entries in each column is TUM (Schrijver, 1986, Truemper, 1977). To show that M^* is TUM, we only need to show that each submatrix M of M^* with at most two nonzero entries in each column is TUM. By a characterization theorem of Camion (Schrijver, 1986), it is sufficient to show that M is balanced, i.e., in every square submatrix M' of M with exactly two nonzero entries per row and per column, the sum of the entries is a multiple of four.

Note each column and each row in M' contains two 1s. So start from any entry and alternate between rows and columns, if it returns to the starting entry, we find a cycle. If not, we continue adding entries and it eventually returns to the starting entry. It implies that M' contains finitely many (entry) disjoint cycles, where each cycle contains a submatrix of M' with exactly two 1s entries per row and per column. We show that each cycle P is even, i.e., the sum of entries in P is a multiple of four.

First note that for any pair of two 1s in each row of P, there are two types of possible changes of column indexes: (i) from (i, j, t) to (i, j', t) in row (i, t) of C_1, and from (i, j, t) to (i', j, t) in row (j, t) of C_2, and (ii) from (i, j, t_i, t_{-i}) to (i, j, t_i, t'_{-i}) in row (i, j, t_i) of B. That is, P induces index changes P_1 on $N \times A$ and P_2 on T. The number of index changes (or rows) in P_1 (i.e., $(i, j) \to (i', j) \to (i', j') \to \cdots \to (i, j)$) is even as there are two hierarchies. The number of index changes in P_2 (i.e., $t_{-i} \to t'_{-i} \to \cdots \to t_{-i}$) is also even since each player has two types. So the number of rows in P is even. This implies that P is an even cycle.

We conclude that each cycle P contains an even number of rows. This completes the proof that M' is balanced, and M (and M^*) is TUM.

(b) General constraints. Pick any square submatrix M of M^*. It is sufficient to prove that $det(M) \in \{0, \pm 1\}$. We claim that after some elementary row operations M can be converted into a matrix $\tilde{M}$ where for each column, there is at most one 1 for each hierarchy. Specifically, for any two rows of C_l $(l = 1, 2)$, (k, t) and (k', t), let G_k and $G_{k'}$ be the corresponding sets in $\mathcal{G}_l$. Suppose $G_k \subset G_{k'}$. Negate row (k, t) and add it to row (k', t). The elementary row operation changes only the sign of the determinant of M and hence $|det(M)| = |det(\tilde{M})|$. It is sufficient to show that $\tilde{M}$ is TUM. Since $\tilde{M}$ now contains at most three 1s in each column, we only need to show that each submatrix of $\tilde{M}$ with at most two nonzero entries in each column is TUM. Following the proof of part (a), we obtain this result.

(3) Suppose each T_i has one element. Then matrix B has only one entry in each row. Since TUM is preserved under deleting unitary columns, we only need to

show that C is TUM. Since $\mathcal{G}$ is the union of two hierarchies, the Edmonds theorem follows the result. Pick any submatrix M of C and we show $|det(M)| \in \{0, \pm 1\}$. By elementary row operations, the submatrix in each hierarchy reduces to a matrix with at most one 1 in each column. The resulting matrix $\hat{M}$ contains at most two 1s in each column and $|det(M)| = |det(\hat{M})|$. If there is some column with no ones, $|det(\hat{M})| = 0$. If there exists a column with a single one, we can expand by minors about that entry and the proof follows by induction. If all columns have two ones, the sums of the rows from two hierarchies must equal and $|det(\hat{M})| = 0$. This completes the proof. □

A.4 Proof of Theorem 3

Proof of Theorem 3. By a well-known characterization theorem on total unimodularity (see Theorem 19.3 (vi), p.269 of Schrijver (1986)), to show that the conic constraint matrix $M^* = \begin{bmatrix} B \\ C \end{bmatrix}$ is TUM, we only need to show that in any square submatrix M of M^* with even row and column sums, the sum of the entries in M is divisible by four. Without loss of generality, assume M has rows from both B and C (without singletons). Notice that if the graph H contains no odd cycle, then H can be represented as a bipartite graph $H = (V_1 \cup V_2, E)$. So for the rows from C in M, each row has exactly two ones. Also note that for the rows from B, each row contains an even number of ones.

We claim that M can be decomposed into finitely many (entry) disjoint cycles, where each cycle has even row and column sums. To see this, start with any entry (i, t_i, t) in M, find an adjacent entry in the same column and alternate between entries in the same rows and columns. If it returns to the starting entry, we find a cycle P: $(i, t_i, t) \to (i, t) \to (i', t) \to \cdots \to (j, t) \to (j, t_j, t) \to (j, t_j, t') \to \cdots \to (k, t_k, t') \to \cdots \to (i, t_i, t'') \to (i, t_i, t)$. Remove all entries of the cycle and repeat. If it does not return to (i, t_i, t), we can continue adding new entries and it eventually returns to (i, t_i, t) since the non-zero entries in M are finite.

We now show that for each cycle P, the sum of the entries in P is divisible by four. First consider the projection P_t of P onto $t \in T$ (i.e., the index changes within each t): $(i, t_i, t) \to (i, t) \to (i', t) \to \cdots \to (j, t) \to (j, t_j, t)$. Then P_t is a path with tail i and head j, i.e., $P_t = P_t^{i,j}$. P_t is called an odd path if i and j are in the same partition V_l; and P_t is an even path if i and j are in the different partitions V_l and

$V_{l'}$. If P_t is odd (even), then the sum of entries in P_t is a multiple of two (four).

Next consider index changes in P by varying $t \in T$: $P_t^{i,j} \to P_{t'}^{j,k} \to \cdots \to P_{t''}^{l,i} \to P_t^{i,j}$. Since each agent has two types, it implies the number of index changes in type profiles is even in P, i.e., the number of P_t is even. Further, it implies both the numbers of odd and even paths are even. Suppose not and the numbers of odd and even paths are odd. By suppressing all odd paths, an odd number of even paths forms an odd cycle in H. But it is a contradiction as H is bipartite. Hence there is an even number of odd paths in P. This shows that the sum of the entries in P is divisible by four. □

A.5 Proof of Theorem 4

Proof of Theorem 4. Substitute the extreme rays in Lemma 7 into the conditions in Lemma 6, we obtain the condition in Theorem 4. □

A.6 Proof of Lemmas 1, 2 and 3

Proof of Lemma 1. Applying Lemma 2 to the implementation system F, we obtain Lemma 1. □

Proof of Lemma 2. We first deal with the 'only if' part. Suppose y is implementable by some x. Pick any $(f, g) \in P$ and $f^\top A^1 + g^\top A^2 = 0$, $g \geq 0$. Then $(f^\top B^1 + g^\top B^2)y \leq -(f^\top A^1 + g^\top A^2)x + f^\top b^1 + g^\top b^2 = f^\top b^1 + g^\top b^2$. Hence $y \in Y$.

Now we turn to the 'if' part. Suppose y is not implementable. There exists no x such that $A^1 x = b^1 - B^1 y$ and $A^2 x \leq b^2 - B^2 y$. By Farkas' Lemma (Schrijver, 1986, p.89), there exists (f, g) such that $f^\top A^1 + g^\top A^2 = 0$, $g \geq 0$, and $f^\top(b^1 - B^1 y) + g^\top(b^2 - B^2 y) < 0$. But then $(f, g) \in P$. There must be $(\hat{f}, \hat{g}) \in \hat{P}$ such that $(\hat{f}^\top B^1 + \hat{g}^\top B^2)y > \hat{f}^\top b^1 + \hat{g}^\top b^2$, and hence $y \notin Y$. □

Lemma 3 is immediately obtained by using the following mathematical result, which gives a nice and clear characterization of general cones defined by a totally unimodular matrix (TUM) and will be derived from a well-known lemma of Hoffman (1976).

We first introduce some notations. Let I be the index set of the inequalities in the cone P. Let $J = I^=(x)$ denote the index set in I for which the corresponding inequalities hold as equations (or active constraints) at $x \in P$. Let M_J be the corresponding submatrix for J, and let rkM_J denote the rank of M_J. Note that $x \in P$ is an extreme ray if and only if $I^{=}(x)$ is maximal, i.e., there exists no $x' \in P$ such that $J' = I^{=}(x')$ and $J \subset J' \subset I$. If x is an extreme ray, then so is λx for all $\lambda \geq 0$. Observe that for a pointed cone, the set of extreme rays provides a unique (up to positive scaling) minimal set of generators.

Lemma 12. *Let $P = \{x \in \mathbb{R}^p \mid Mx \leq 0\}$ be a polyhedral cone and let M be TUM.*

(1) If P is pointed, then P is generated by $\{0, \pm 1\}$ extreme rays.

(2) If P is non-pointed, then P is generated by $\{0, \pm 1\}$ vectors.

Proof of Lemma 12. (1) Let P be pointed. Assume that $z \in P$ is an extreme ray and $J = I^{=}(z)$. Since rk$M_J = q - 1$, there is a submatrix M' with $q - 1$ linearly independent rows in M_J. Since M is TUM, M' is also TUM. We need to show that if M' with columns $M'_1, ..., M'_q$ has rank $q - 1$, then $M'z = 0$ implies that all nonzero coordinates of z are either α or $-\alpha$, for some $\alpha > 0$. The proof is essentially the same as Lemma 3.1 of Hoffman (1976). Note that for any $z_j \neq 0$,

$$M'_j = \sum_{i=1, i \neq j}^{q} \frac{z_i}{z_j} M'_i,$$

and M'_i, $i \neq j$ are linearly independent. The linear system has a unique solution $(\frac{z_i}{z_j}), i \neq j$. Since M' is TUM, each $\frac{z_i}{z_j}$ is integer. As this argument applies to every nonzero entry in z, for any nonzero $z_j, z_{j'}$ and $j \neq j'$, $\frac{z_{j'}}{z_j}$ and $\frac{z_j}{z_{j'}}$ are integers and hence $|z_j| = |z_{j'}| = \alpha$ for some $\alpha > 0$. Therefore every extreme ray of P contains a $\{0, \pm 1\}$ vector.

(2) Let P be non-pointed. We write P as a union of finitely many pointed cones. Notice that the Euclidean space $\mathbb{R}^q$ is a union of $l := 2^q$ closed orthants $i = 1, ..., l$. Let P_i be the intersection of P and orthant i (with some P_i possibly empty). Then $P = \cup_{i=1}^{l} P_i$. We claim that each P_i and hence P is generated by $\{0, \pm 1\}$ vectors. By the first part of the lemma, we only need to show that for each i, P_i is a pointed cone and its constraint matrix M_i is TUM. Notice that by construction, each P_i is a pointed cone. Since M is TUM and M_i is obtained from M by adding rows with at most one non-zero entries (1 or -1), M_i is also TUM. □

A.7 Proof of Lemma 6

We provide a detailed proof for Lemma 6. Denote $\mathbf{C}_2 \subset \mathbf{C}^T \times \mathbf{C}^T$ the set of all possible pairs of profile of chains on disjoint ground sets, i.e., if $(\mathcal{C}_1, \mathcal{C}_2) \in \mathbf{C}_2$, then for each $t \in T$, $\mathcal{C}_1^t$ and $\mathcal{C}_2^t$ have disjoint ground sets.

Lemma 13. *Suppose pair (b, c) is paramodular. Then the following two systems on $y \in \mathbb{R}^l$ are equivalent:*

System F_1: $f^\top y \le g^\top c - h^\top b$, for all $(f, g, h) \in W$, and

System F_2: $f^\top y \le g^\top c - h^\top b$, for all $(f, g, h) \in \cup_{(\mathcal{C}_1, \mathcal{C}_2) \in \mathbf{C}_2} W(\mathcal{C}_1, \mathcal{C}_2)$.

Proof of Lemma 13. First notice that $\cup_{(\mathcal{C}_1, \mathcal{C}_2) \in \mathbf{C}_2} W(\mathcal{C}_1, \mathcal{C}_2) \subset W$ and hence F_2 is a subsystem of F_1. We show that the inequalities in F_2 implies all the other inequalities in F_1. By Lemma 5, for any $f \in \mathbb{R}^l$, the tightest upper bound for $f^\top y$ is attained at some (g, h) where for each $t \in T$, (g_t, h_t) have supports given by some chains $(\mathcal{C}_1^t, \mathcal{C}_2^t)$ on disjoint ground sets (in $N \times A$). We have $(f, g, h) \in \cup_{(\mathcal{C}_1, \mathcal{C}_2) \in \mathbf{C}_2} W(\mathcal{C}_1, \mathcal{C}_2)$. Hence for any $f \in \mathbb{R}^l$, the inequalities in F_2 gives the tightest bound and imply all the other inequalities in F_1. $\square$

We next show that system F_2 in Lemma 13 is equivalent to the system in Lemma 6. That is, we can split the condition in F_2 into two separate conditions, which correspond to ceiling and floor constraints independently.

Proof of Lemma 6. For any $f \in \mathbb{R}^l$, suppose the tightest upper bound for $f^\top y$ is attained at some (g, h) whose supports are some chains $\mathcal{C}_1$ and $\mathcal{C}_2$ on disjoint ground sets. For each $t \in T$, let f_t denote the projection of f onto t, i.e., $f_t(i, j) = f(i, t_i, j)$. Define f_t^+ the coordinates of f_t where $f_t(i, j) > 0$ and f_t^- the coordinates of f_t where $f_t(i, j) < 0$. Then $f_t(i, j) > 0$ implies $(i, j) \in \texttt{supp}(g_t)$ and $(i, j) \notin \texttt{supp}(h_t)$. It implies $-f_t^+ + \sum_{U \in \mathcal{C}_1^t} g_t(U) \chi^U = 0$. Hence $(f^+, g) \in W^+(\mathcal{C}_1)$. Similarly, we have $(f^-, h) \in W^-(\mathcal{C}_2)$. Then the inequalities in the sets $\cup_{\mathcal{C}_1 \in \mathbf{C}^T} W^+(\mathcal{C}_1)$ and $\cup_{\mathcal{C}_2 \in \mathbf{C}^T} W^-(\mathcal{C}_2)$ imply all inequalities in $\cup_{(\mathcal{C}_1, \mathcal{C}_2) \in \mathbf{C}_2} W(\mathcal{C}_1, \mathcal{C}_2)$. System F_2 in Lemma 13 is equivalent to the system in Lemma 6. $\square$

A.8 Proof of Lemma 7

Proof of Lemma 7. We prove the result for $W^+(\mathcal{C}_1)$ ($W^-(\mathcal{C}_2)$ can be shown analogously). (Only if) Suppose (f, g) is an extreme ray. By Lemma 6, (f, g) is 0-1 valued.

It implies that f is the incidence vector of some S. Then g satisfies

$$\sum_{U \in \mathcal{C}_1^t} g_t(U)\chi^U = \chi^{I(t,S)}, \tag{35}$$

$$g_t(U) \in \{0, 1\}. \tag{36}$$

Since each $g_t(U) \in \{0, 1\}$, g_t is the incidence vector of some $\mathcal{P} \subseteq \mathcal{C}_1^t$: $g_t(U) = 1$ for each $U \in \mathcal{P}$ and $g_t(U) = 0$ otherwise. It implies that each (i, j) in $I(t, S)$ is contained in one and only one $U \in \mathcal{P}$. Since $\mathcal{P}$ is a chain, $\mathcal{P}$ must contain a single element, i.e., $\mathcal{P} = \{I(t, S)\}$.

(If) Suppose (f, g) given in the Lemma is not an extreme ray, i.e., the set of active constraints in (35)-(36) is not maximal. But then there exists $(\tilde{f}, \tilde{g})$ such that more constraints in (36) are active. But then $\tilde{g}_t(U) = 0$ for all U. Contradiction. □

A.9 Proof of Lemma 8

Proof of Lemma 8. Since for each family $i \in Y$, only two cars of different models are acceptable, it implies that $0 \leq x(i, j) \leq 1$, for all $j \in A$. Hence each demand constraint $i \in X \cup Y$ is canonical, and (22)-(24) form one hierarchy. Combine with the supply constraints we have (21)-(24) are bihierarchical. □

A.10 Proof of Theorem 5

Proof of Theorem 5. The theorem immediately follows from Corollary 1 when the floor constraint is standard the general capacity function is linear. □

A.11 Proof of Lemmas 9, 10 and 11

Proof of Lemma 9. (1) $(f, g)M(W) = 0$ implies $f(i, t_i) = g(t) = f(j, t_j) = g(t'_i, t_j) = f(i, t'_i)$ for all t_i, t_j, t'_i. Hence $(f, g) = (1, ..., 1)$ is in the linearity space of W and W is non-pointed. (2) $M(W)$ is TUM since by deleting unitary column we obtain the node-arc incidence matrix of $(T_1 \cup T_2, E)$. □

Proof of Lemma 10. Let (f, g) be a $\{0, \pm 1\}$ generator. Since $(f, g) \in W$, for each $t \in E$, we have

$$-f(i, t_i) + g(t) \geq 0. \tag{37}$$

When all of the constraints are active, the generators in the linearity space of W is given by $(1, ..., 1)$ and $(-1, ..., -1)$. We next identify the generators not in the linearity space. We consider the following two cases.

(a) $f(i, t_i) \in \{0, -1\}$ for all (i, t_i). Since (f, g) attains a maximal set of active constraints, then $f(i, t_i) = -1$ for a unique (i, t_i) and 0 otherwise, and $g = 0$.

(b) $f(i, t_i) = 1$ for at least one (i, t_i). By (37), for any $t \in E$, if $f(i, t_i) = 1$ for at least one i, then $g(t) = 1$. If $f(i, t_i) = 0$ for exactly one i, and $f(k, t_k) \leq 0$, then $g(t) = 0$. If $f(i, t_i) = -1$ for $i = 1, 2$, then $g(t) = -1$. Define $(\tilde{f}, \tilde{g})$ by replacing each $f(i, t_i) = -1$ by 0 and $g(t) = -1$ by 0. There are more active constraints at $(\tilde{f}, \tilde{g})$ than at (f, g). Hence if (f, g) has a maximal set of active constraints, $f \geq 0$ and (f, g) is the incidence vector of $(S, \Gamma(S))$. □

We introduce the following theorem for the proof of Lemma 11.

Lemma 14. *(Ghouila-Houri, 1962; Schrijver, 1986) Let M be an $p \times q$ matrix. M is totally unimodular if and only if for every subset of columns $\Omega \subseteq \{1, ..., q\}$, there exists a partition Ω_1, Ω_2 of Ω such that*

$$\left| \sum_{j \in \Omega_1} m_{ij} - \sum_{j \in \Omega_2} m_{ij} \right| \leq 1 \;\; for \;\; i = 1, \ldots, p.$$

Proof of Lemma 11. We show that the constraint submatrix $M = \begin{bmatrix} B \\ C \end{bmatrix}$ is totally unimodular. Note that M (without singletons) is given by

$$M = \begin{bmatrix}
1 & 1 & & & & & & \\
& & 1 & 1 & & & & \\
& & & & 1 & 1 & & \\
& & & & & & 1 & 1 \\
1 & & 1 & & & & & \\
& 1 & & 1 & & & & \\
& & & & 1 & & 1 & \\
& & & & & 1 & & 1 \\
1 & & & & 1 & & & \\
& 1 & & & & 1 & & \\
& & 1 & & & & 1 & \\
& & & 1 & & & & 1
\end{bmatrix}.$$

Partition all columns of M into

$$\Omega_1 = \{(a_1, t_{11}), (a_1, t_{22}), (a_2, t_{12}), (a_2, t_{21})\} \text{ and}$$
$$\Omega_2 = \{(a_1, t_{12}), (a_1, t_{21}), (a_2, t_{11}), (a_2, t_{22})\},$$

where t_{kl} denotes the type profile (t_{1k}, t_{2l}). For any subset Ω of all columns, let

$$\Omega_1' = \Omega \bigcap \Omega_1 \text{ and } \Omega_2' = \Omega \bigcap \Omega_2.$$

It can be seen that the two 1s in each row either belongs to different sets Ω_1' and Ω_2', or at least one of the two 1s belongs to neither Ω_1' or Ω_2'. By the Ghouila-Houri theorem, M is TUM. □

References

Alaei, S., Fu, H., Haghpanah, N., Hartline, J., and Malekian, A. (2019). Efficient computation of optimal auctions via reduced forms. *Mathematics of Operations Research*, 44(3):1058–1086.

Arrow, K. and Hahn, F. (1971). *General Competitive Analysis*. North-Holland.

Balas, E. (2001). Projection and lifting in combinatorial optimization. *Computational Combinatorial Optimization: Optimal or Provably Near-Optimal Solutions, Juenger, M. and Naddef, D., eds. Springer.*

Balas, E. and Pulleyblank, W. (1983). The perfectly matchable subgraph polytope of a bipartite graph. *Networks*, 13:495–516.

Biró, P., Fleiner, T., Irving, R., and Manlove, D. (2010). The college admissions problem with lower and common quotas. *Theoretical Computer Science*, 411:3136–3153.

Bogomolnaia, A. and Moulin, H. (2001). A new solution to the random assignment problem. *Journal of Economic Theory*, 100(2):295–328.

Bondy, J. and Murty, U. (1976). *Graph Theory with Applications*. MacMillan.

Border, K. (1991). Implementation of reduced form auctions: A geometric approach. *Econometrica*, 59(4):1175–1187.

Border, K. (2007). Reduced form auctions revisited. *Economic Theory*, 31:167–181.

Börgers, T. and Postl, P. (2009). Efficient compromising. *Journal of Economic Theory*, 144(5):2057–2076.

Budish, E. and Cantillon, E. (2014). The multi-unit assignment problem: Theory and evidence from course allocation at harvard. *American Economic Review*, 102(5):2237–2271.

Budish, E., Che, Y., Kojima, F., and Milgrom, P. (2013). Designing random allocation mechanisms: Theory and applications. *American Economic Review*, 103(2):585–623.

Cai, Y., Daskalakis, C., and Weinberg, M. (2012). Optimal multi-dimensional mechanism design: Reducing revenue to welfare maximization. *Proceedings of 53rd Annual IEEE Symposium on Foundations of Computer Science*, pages 130–139.

Che, Y., Kim, J., and Mierendorff, K. (2013). Generalized reduced-form auctions: A network-flow approach. *Econometrica*, 81(6):2487–2520.

Cheng, Y. and Yang, Z. (2017). A stable hospital-doctor matching mechanism under distributional and hierarchical constraints. *Preprint*, University of York.

Crawford, V. and Knoer, E. (1981). Job matching with heterogeneous firms and workers. *Econometrica*, 49(1):437–450.

Debreu, G. (1959). *Theory of Value*. Yale University Press.

Demange, D., Gale, D., and Sotomayor, M. (1986). Multi-item auctions. *Journal of Political Economy*, 94(1):863–872.

Demange, G. (2004). On group stability in hierarchies and networks. *Journal of Political Economy*, 112:754–778.

Edmonds, J. (1970). Submodular functions, matroids, and certain polyhedra. *Combinatorial Structures and Their Applications*, Gordon and Breach.

Ehlers, L., Hafalir, I., Yenmez, B., and Yildirim, M. (2014). School choice with controlled choice constraints: Hard bounds versus soft bounds. *Journal of Economic Theory*, 153:648–683.

Fragiadakis, D. and Troyan, P. (2016). Improving matching under hard distributional constraints. *Theoretical Economics*, 12(2):863–908.

Frank, A., Király, T., Pap, J., and Pritchard, D. (2013). Characterizing and recognizing generalized polymatroids. *Mathematical Programming, Series A*.

Fujishige, S. (2005). *Submodular Functions and Optimization, 2nd edition*. Elsevier.

Gershkov, A., Goeree, J., Kushnir, A., Moldovanu, B., and Shi, X. (2013). On the equivalence of bayesian and dominant strategy implementation. *Econometrica*, 81(6):197–220.

Goeree, J. and Kushnir, A. (2016). Reduced-form implementation for environments with value interdependencies. *Games and Economic Behavior*, 99:250–256.

Goeree, J. and Kushnir, A. (2022). A geometric approach to mechanism design. *Journal of Political Economy: Microeconomics*.

Gopalan, P., Nisan, N., and Roughgarden, T. (2015). Public projects, boolean functions and the borders of border's theorem. *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, pages 395–415.

Greenberg, J. and Weber, S. (1986). Strong tiebout equilibrium under restricted preferences domain. *Journal of Economic Theory*, 38(1):101–117.

Hart, S. and Reny, P. (2015). Implementation of reduced form mechanisms: A simple approach and a new characterization. *Economic Theory Bulletin*, 3(1-8).

Herings, P., van der Laan, G., Talman, A., and Yang, Z. (2010). The average tree solution for cooperative games with communication structure. *Games and Economic Behavior*, 68:626–633.

Heydenreich, B., Müller, R., Uetz, M., and Vohra, R. (2009). Characterization of revenue equivalence. *Econometrica*, 77(1):307–316.

Hoffman, A. (1976). Total unimodularity and combinatorial theorems. *Linear Algebra and Applications*, 13(1-2):103–108.

Kamada, Y. and Kojima, F. (2015). Efficient matching under distributional constraints: Theory and applications. *American Economic Review*, 105(1):67–99.

Kelso, A. and Crawford, V. (1982). Job matching, coalition formation, and gross substitutes. *Econometrica*, 50:1483–1504.

Koopmans, T. and Beckmann, M. (1957). Assignment problems and the location of economic activities. *Econometrica*, 25(1):53–76.

Lang, X. and Mishra, D. (2023). Symmetric reduced form voting. *Theoretical Economics*, forthcoming.

Lang, X. and Yang, Z. (2021). Reduced-form allocations for multiple indivisible objects under constraints. *DERS Working Paper No. 21/04 and 21/05, University of York.*

Manelli, A. and Vincent, D. (2010). Bayesian and dominant-strategy implementation in the independent private values model. *Econometrica*, 78(6):1905–1938.

Maskin, E. and Riley, J. (1984). Optimal auctions with risk averse buyers. *Econometrica*, 52(6):1473–1518.

Matthews, S. (1984). On the implementability of reduced form auctions. *Econometrica*, 52(6):1519–1522.

Milgrom, P. and Segal, I. (2020). Clock auctions and radio spectrum reallocation. *Journal of Political Economy*, 128:1–31.

Myerson, R. (1977). Graphs and cooperation in games. *Mathematics of Operations Research*, 2:225–229.

Myerson, R. (1981). Optimal auction design. *Mathematics of Operations Research*, 6(1):58–73.

Myerson, R. and Satterthwaite, M. (1983). Efficient mechanisms and bilateral trading. *Journal of Economic Theory*, 29(2):265–281.

Rothkopf, M., Pekec, A., and Harstad, R. (1998). Computationally manageable combinatorial auctions. *Management Science*, 44(8):1131–1147.

Schrijver, A. (1986). *The Theory of Linear and Integer Programming.* John Wiley and Sons.

Seymour, P. (1980). Decomposition of regular matroids. *Journal of Combinatorial Theory (B)*, 28(1):305–359.

Seymour, P. (1981). Recognizing graphic matroids. *Combinatorica*, 1(1):75–78.

Shapley, L. and Shubik, M. (1971). The assignment game i: The core. *International Journal of Game Theory*, 1(1):111–130.

Sönmez, T. and Ünver, U. (2010). Course bidding at business school. *International Economic Review*, 51(1):99–123.

Truemper, K. (1977). Unimodular matrices of flow problems with additional constraints. *Networks*, 7:343–358.

Vohra, R. (2013). Lecture notes in submodularity and discrete convexity. *University of Pennsylvania*.

Williamson, O. E. (1975). *Markets and Hierarchies: Analysis and Antitrust Implications*. The Free Press.

Zheng, C. (2021). Reduced-form auctions of multiple objects. *Unpublished working paper, University of Western Ontario.*