A characterization of the existence of succinct linear representation of subset-valuations

Saša Pekeč*

The Fuqua School of Business

Duke University

Abstract

Decisions that involve bundling or unbundling a large number of objects, such as deciding on the bundle structure or optimizing bundle prices, are based on underlying valuation function over the set of all possible bundles. Given that the number of possible bundles (i.e., subsets of the given set of objects) is exponential in the number of objects, it is important for the decision-maker to be able to represent this valuation function succinctly. Identifying all structural sources of synergy in subset valuations might point to simple and concise representation of the valuation function. We characterize additive and multiplicative representations of synergies in subset valuations and subset utility, which in turn points to necessary and sufficient conditions for a succinct representation of subset valuations to exist.

*This work is to honor Peter Fishburn, whose contribution and guidance was critical to formulating and proving the main result. I am grateful for Peter's friendship and mentorship, and am honored to have had opportunities to collaborate with him.

1 Introduction

Representation of valuations over multiple objects is foundational in modeling, from models of human behavior and choice to models of economic and market mechanisms. These valuations are often interrelated, and if modeling involves choice or decisions over sets of objects, an additional information for representing valuations over sets of objects may be necessary. This work provides a characterization for the existence of succinct linear representation of valuations over sets of objects, thereby indicating what structural properties are needed to ensure practical tractability of such representation, and suggesting what information would need to be elicited.

The issue of selecting or choosing a subset of available multiple objects arises naturally in many practical settings. For example, understanding a customer's valuations over sets of objects is critical for an e-ratailer's personalized assortment and pricing decision tailored for that customer (e.g., see [9, 26]). Similarly, a seller of customizable product, from a software service to a car, needs to understand customer valuations over sets of additional options or features in order to decide on optimal bundling (e.g., see [25, 3, 36]). For another example, consider digital advertising markets that in real time insert targeted ads into a content shown to users. The advertiser valuation for a user depends on the user's attributes (e.g., location, demographics, etc.) which are represented by the user's digital imprint. This valuation is not simply an additive function, e.g., the value for a female older than 65 in New York City need not be equal to the sum of values the advertiser has for reaching a female, a New York City resident, and a person older than 65. In fact, digital advertising markets are designed to allow advertisers express their valuations over sets of user attributes and to match each user with an advertiser in real-time by embedding the ad into the content as it is being loaded by the user (e.g., see, [2, 5, 8, 10]). Therefore, understanding preferences over subsets is a prerequisite for making meaningful decisions in creating, choosing, allocating or pricing multiple objects.

The fundamental problem of eliciting or estimating preferences over the set of all subsets is that the number of subsets could be prohibitively large irrespective of computational and communicational power of the decision-maker. For example, more than a quadrillion possible subsets can be created from 50 objects (since $2^{50} - 1 > 1.1$ quadrillion). Thus, it is not surprising that even the problem of choosing a subset of a finite set is known to be computationally unmanageable, [35]. This intractability in developing or utilizing fully general models for subset valuations to guide subset choice, allocation and/or pricing decisions, underscores the need for identifying settings in which a succinct, tractable representation of subset valuations is possible.

In the aforementioned assortment problem, prevailing practice has been to employ data-driven approaches and statistical estimation techniques (see, e.g., [24, 6]), typically assuming that the assortment value is simply a linear function of the values of its individual objects. This ignores substitutabilities and complementarities among objects in the assortment and hence could yield suboptimal decisions. Some recent work [4], however, has extended a classical multinomial logit (MNL) model from individual item selection to subset selection, by expanding this linear valuation with additional terms corresponding to capturing synergetic impacts for a small number of subsets. Practical implementability and usefulness of this approach hinges on the ability to identify subsets and/or features driving these adjustments.

The challenge of modeling subset valuations is critical for combinatorial auctions [12]: market mechanisms in which sets of objects are traded. In combinatorial auctions bidders submit bids on bundles, i.e., subsets of the set of objects (instead of having to submit separate bids on single objects). Thus, every bidder needs to understand its own preferences over all possible bundles. On the other hand, the bid-taker has to choose a combinatorial auction procedure that elicits enough information about bidders' bundle valuations, so that an optimal allocation and pricing decision can be made. Thus, an efficient elicitation of bundle valuations is a central issue of combinatorial auction design [31] that has stimulated a considerable research interest (e.g., [11, 29, 30, 33, 8]). For example, one of the approaches in the design and use of combinatorial auctions is to approximate bundle valuations by a simple function. The most natural candidate is the linear approximation (a.k.a. additive subset utility) that estimates the value of any bundle by the sum of the values of the individual objects in that bundle. The number of parameters in this case equals the number of objects, which is considerably smaller than the number of all bundles. The problem with such a simple approximation is that it neglects synergetic effects of bundle valuations. After all, the reason for considering creation, choice, allocation and/or pricing of bundles is to exploit potential synergies from the bundling process. However, it could be true that potential sources of synergies are limited and easily identified (several examples can be found in [32]), in which case bundle valuations might be represented in a succinct and simple way, yielding a readily implementable combinatorial auction that does not suffer from computational intractability.

This note provides a necessary and sufficient condition for subset valuations (and subset utility) to be representable as a sum (or a product) of different synergetic effects. (For example, an advertiser who values reaching out to females or New York City residents or those older than 65, might put an additional value for reaching a user who has all three of these features.) More precisely, given a collection of properties that each subset could have or not have, we characterize the existence of weights, one for each such property, such that the utility of any subset is the sum of the weights corresponding to the properties that subset possesses. One special case, where only sources of synergy are binary interactions (i.e., synergetic effects are limited to pairs of objects that a subset contains) has been the topic of [22]. The results presented here generalize those in [22]. However, our main point is not that results from [22] can be generalized, but it is that our generalization indicates that correctly identifying sources of synergy could yield to simple subset valuation and subset utility functions.

2 Representation

Denote the finite set of indivisible items by $[n] = \{1, 2, \ldots, n\}$. The set of all possible subsets that could be constructed from these n items, i.e., the set of all subsets of $[n]$ is denoted by $2^{[n]}$. We will sometimes abuse notation and denote the singleton set $\{x\}$ by x.

We assume that every two subsets $A, B \in 2^{[n]}$ can be compared. We use $\succ$ to denote the strict preference relation on $2^{[n]}$, so that $A \succ B$ when A is preferred to B. We say that $\succ$ is *complete* if, for any two sets A and B, $A \neq B$, either $A \succ B$ or $B \succ A$. If neither $A \succ B$ nor $B \succ A$, we write $A \sim B$. In other words, $\sim$ is the indifference relation derived from $\succ$. We use $A \succeq B$ to denote that B is not preferred to A, i.e., that either $A \succ B$ or $A \sim B$.

The preference relation $\succ$ is *asymmetric* if for every A, B, $A \succ B$ implies that not $B \succ A$.

The preference-indifference relation $\succeq$ is *transitive* if, for every A, B, C, $A \succeq B$ and $B \succeq C$ imply $A \succeq C$. As in [14], $\succ$ is a *weak order* if and only if (i) $\succ$ is asymmetric and (ii) $\succeq$ is transitive.

Throughout, we make the following

Assumption. *The preference relation $\succ$ is a weak order.*

If a preference relation over subsets of $[n]$ were not a weak order, one would think that there is some inconsistency, since both asymmetry and transitivity are quite natural conditions in this setting.

Proposition 1 *There exists $u : 2^{[n]} \to \Re$, $u(\emptyset) = 0$, such that*

$$A \succ B \Leftrightarrow u(A) > u(B). \tag{1}$$

The fact that a weak order over a finite set can have a representation in real numbers is well-known, e.g., see [14].

Note that the representation u from Proposition 1 is not unique: for any strictly increasing $f : \Re \to \Re$ with $f(0) = 0$, $f \circ u$ is also a representation of $\succ$. When does a simple representation u of $\succ$ exist? In particular, we are interested in functions u that are linear functions with a manageable number of terms.

Example 2 An additive subset utility.

Suppose $u(A \cup B) = u(A) + u(B) - u(A \cap B)$. Then $u(S) = \sum_{x \in S} u(x)$ for all $S \subseteq [n]$. Thus, u is an additive function completely defined by the n values $u(x)$ for $x \in [n]$. $\blacksquare$

Next, we define a *set property* to be any function $P : 2^{[n]} \to \{0,1\}$. P can be viewed as a characteristic function of a set property: if $P(A) = 1$, then set A has property P, and if $P(A) = 0$ then set A does not have property P. Here are some simple examples:

- One could be interested in sets of a certain size, since economies of scale could be important. Thus, one can define property $P_{\geq k}$ by setting $P_{\geq k}(A) = 1$ if and only if $|A| \geq k$. Obviously, similar cardinality-based set properties can also be defined (e.g., $P_{\leq k}$, $P_{=k}$).

- Set properties could take into account synergetic values that come out of topological properties of the underlying set of items. For example, if items are vertices of a graph, one can define set functions corresponding to many graph properties, e.g., set $P_{conn}(A) = 1$ if and only if the subgraph induced by A is connected. For another graph example, if items are edges of some graph, one could, e.g., set $P_{ST}(A) = 1$ if and only if A contains a spanning tree of the graph.

- For every $S \subset [n]$, property P_S can describe whether set A contains S or not. Set $P_S(A) = 1$ if and only if $S \subseteq A$. A special case occurs when $S = \{x\}$, i.e., setting $P_x(A) = 1$ if and only if $x \in A$. Note that an additive subset utility u from Example 2 is a linear combination of P_x functions:

$$u = \sum_{x \in [n]} u(x) P_x.$$

Further, additionally setting $P_{xy}(A) = 1$ if and only if $\{x, y\} \subseteq A$, allows for going beyond individual elements, and capturing contributions of pairwise interactions to

$u(S)$:

$$u = \sum_{x \in [n]} w_x P_x + \sum_{\{x,y\} \subseteq [n]} w_{xy} P_{xy},$$

where $w_x = u(x)$ and $w_{xy} = u(\{x, y\}) - u(x) - u(y)$. Note that this is the functional format of subset valuations whose existence is characterized in [22].

In fact, any $\succ$ can be represented by a utility function that is a linear combination of set property functions:

Proposition 3 *Every representation u of any weak order $\succ$ on $2^{[n]}$ is a linear combination of at most $2^n - 1$ set property functions. The set of representations $u \in R^{2^n-1}$ of a complete $\succ$ that can be decomposed into a linear combination of less than $2^n - 1$ set property functions has measure zero.*

Proof. For every nonempty set $A \subseteq [n]$, define its indicator set function P_A by setting $P_A(S) = 1$ if and only if $S = A$. Then

$$u = \sum_{A \subseteq [n], A \neq \emptyset} u(A) P_A.$$

We now show that the set of representations u of a complete $\succ$ that can be decomposed into a linear combination of less than $2^n - 1$ set property functions, has measure zero. Let $P_1, \ldots P_k$ be set property functions and $w_1, \ldots w_k$ real numbers such that

$$u = \sum_{j=1}^{k} w_j P_j.$$

Then, denoting all nonempty subsets of $[n]$ by $A_1, A_2, \dots, A_{2^n-1}$, we have

$$u(A_i) = \sum_{j=1}^{k} w_j P_j(A_i), \qquad i = 1, \dots 2^n - 1. \tag{2}$$

Let M be a 0-1 matrix with $2^n - 1$ rows and k columns defined by $M_{ij} = P_j(A_i)$. Note that any collection of set properties generates some 0-1 matrix M. Denoting
$\mathbf{u} = (u(A_1), \dots, u(A_{2^n-1}))^T$ and $\mathbf{w} = (w_1, \dots, w_k)^T$, (2) becomes

$$\mathbf{u} = M\mathbf{w}.$$

The rank of M is at most k, and thus, for $k < 2^n - 1$, the image of M is a proper linear subspace of $\Re^{2^n-1}$, which has measure zero. Since there is a finite number of 0-1 matrices with $2^n - 1$ rows and $k < 2^n - 1$ columns, the union of the images of all such matrices has measure zero in $\Re^{2^n-1}$. Thus, the set of representations u that cannot be decomposed into a linear combination of less than $2^n - 1$ set property functions has measure one. ▮

Proposition 3 establishes that a linear representation of subset valuations might generically require exponentially many set functions. This suggests that any method for representing, eliciting or estimating subset valuations in a generic setting will be computationally intractable (pointing to the limitation of the MNL generalization presented in [4], and inherent limitations of other approaches; e.g., [1] establish a more formal connection between different approaches used in machine learning, while [13] make a similar connection between Mallows-based model and certain class of size-independent subset choice models).

However, in many practical settings, there might be a limited number of set properties that determine all subset valuations. In such cases, methods that focus on determining or estimating how these properties contribute to a subset value, could be tractable and implementable.

Example 4 Consider an advertiser with value v for reaching a female New York City resident older than 65, while having a zero value for reaching any users that do not have all of these attributes. User's digital imprint can be thought of as a binary string of length n (e.g., cookie info), so any user is characterized by $S \subseteq [n]$ where $i \in S$ if and only if i^{th} digit of the binary string is 1. Suppose that (female indicator:) f^{th} digit is 1 if and only if the user is identified as female; (location indicator:) one of digits in $L \subset [n]$ is 1 if and only if the user resides in one of the New York City zipcodes; and (age indicator:) one of digits in $A \subset [n]$ is 1 if the user's age is one encoded by that digit. Then, the set property P_{target} describing users that are worth targeting, i.e., valuable to the advertisers, is defined as:

$$P_{\text{target}}(S) = P_f(S) \wedge P_{\geq 1}(S \cap L) \wedge P_{\geq 1}(S \cap A).$$

Then the advertisers valuation of a user $S \subseteq [n]$ is simply

$$u = vP_{\text{target}}.$$

Note that this valuation function format requires determining only a single parameter v, even though there are $1 + |L| + |A|$ elements of $[n]$ that need to be considered when determining whether S has a non-zero value. Further, the advertiser's valuation cannot be decomposed into a sum of values of individual elements (as in additive subset utility) nor a sum that involves values of pairwise interactions (as in the model studied in [22]). An alternative $(|L| \times |A|)$-dimensional representation would require an interaction term for each tripleton $\{f, l, a\}$, $l \in L$, $a \in A$, of features.

Also note that the advertiser's valuation can be expanded with additional terms to capture target users of possibly different value. For example, if the advertiser has value w for female New York City residents in the 55-65 age group, one can analogously define $P_{\text{target2}}(S) = P_f(S) \times P_{\geq 1}(S \cap L) \times P_{\geq 1}(S \cap B)$, with B corresponding to digits representing ages falling in the 55-65 age group. Then, the advertiser's valuation has two terms: $u = vP_{\text{target}} + wP_{\text{target2}}$.

Finally, note that the advertiser subset valuations need to be represented in the setting where each digit of the user's binary string is an object of the set $[n]$. In order to effectively participate in the advertising market, both the advertiser and the market operator (e.g., Google Ad Exchange) need to allow for a computationally tractable representation of these subset valuations. In other words, set properties driving the advertiser's valuation are known to the advertiser, so are exogenous. A difficulty the advertiser faces is whether their subset valuations can be represented using predefined objects of $[n]$ in a way that makes expressing these valuations (and consequent choices and decisions) manageable; e.g., if n is large (which is the case with user data in digital ad markets) it would be unmanageable if the advertiser would need to compute or communicate exponentially many bits of information. ∎

The research question addressed in this paper is that of characterizing $\succ$ which are linear combinations of at most k set properties (given exogenously).

We first define

Cancellation Condition. *A preference relation* $\succ$ *satisfies the cancellation condition for the set properties* $P_1, \ldots, P_k$ *if for any positive integer* j *and any* $A_1, A_2, ..., A_j$ *and* $B_1, B_2, ..., B_j$ *satisfying*

$$|\{l : 1 \leq l \leq j, P_i(A_l) = 1\}| = |\{l : 1 \leq l \leq j, P_i(B_l) = 1\}| \quad \text{for all} \quad i = 1, \ldots, k \tag{3}$$

$A_1 \succ B_1$ *implies that there exists* $m \leq j$ *such that* $B_m \succ A_m$.

If $P_1, \ldots, P_k$ are the only set properties relevant to determining preference among sets that are being compared, then the cancellation condition seems to be a natural requirement: if both collections contain the same number of sets with set property P_i and if that is true for every P_i, then no collection "dominates" the other, i.e., if there is a pair $A_j \succ B_j$, then there has to be another pair $B_{j'} \succ A_{j'}$.

Theorem 5 *A preference relation* $\succ$ *satisfies the cancellation condition for the set properties* $P_1, \ldots, P_k$ *if and only if there exist real numbers* w_i *such that*

$$u = \sum_{i=1}^{k} w_i P_i \tag{4}$$

is a representation of $\succ$.

Proof. Suppose that the cancellation condition holds. We use linear algebra and a theorem of the alternative to show the existence of the representation (4). Let

$$P(A) = (P_1(A), P_2(A), \ldots, P_k(A)).$$

Note that there are $m = 2^{n-1}(2^n - 1)$ pairs of different subsets $\{A, B\}$ and we can without loss of generality assume that $A \succeq B$. If a representation of the form (4) exists, then by (1) and using notation $\mathbf{w} = (w_1, \ldots, w_k)^T$, for each such pair

$$A \succ B \quad \Rightarrow \quad (P(A) - P(B))\mathbf{w} > 0 \tag{5}$$

and

$$A \sim B \quad \Rightarrow \quad (P(A) - P(B))\mathbf{w} = 0. \tag{6}$$

Thus, representation (4) exists if and only if the system of m linear (in)equalities of the form (5) or (6) has a solution $\mathbf{w}$.

Let $\mathbf{M}$ be the $m \times k$ matrix whose rows are vectors $(P(A) - P(B))$. Note that all entries of M are either -1, 0, or 1. The theorem of the alternative (or Farkas Lemma; e.g., see Theorem B in [15] or Corollary 7.1 in [34]) states that either a solution $\mathbf{w}$ exists or there exists a non-negative integral solution of

$$\mathbf{yM} = \mathbf{0} \tag{7}$$

with $y_{i^*} > 0$ for at least one i^* such that the i^*-th row of M corresponds to the inequality (5). Without loss of generality we may assume that $i^* = 1$.

We will complete the sufficiency part of the proof by showing that, if the cancellation condition holds, such solution $\mathbf{y}$ to (7) does not exist, thereby proving that representation (4) exists. Specifically, we will use $\mathbf{y} = (y_1, \ldots, y_m)$ to guide construction of sets $A_1, \ldots, A_j$ and $B_1, \ldots B_j$ on which we will apply the cancellation condition: for every $y_r \neq 0$ we will add y_r identical pairs of sets A, B, that correspond to the (in)equality defining r-th row of M. Thus, the family will have $j = y_1 + y_2 + \ldots + y_m$ pairs of sets A, B, with exactly y_r copies of sets A, B defining r-th row of M (and if $y_r = 0$ there will be no copies). Formally, let $\mathbf{y} = (y_1, \ldots, y_m)$ be a non-negative integral nonzero solution of (7). Construct $A_1, \ldots, A_j$ and $B_1, \ldots B_j$, on which we will apply the cancellation condition, as follows: let $j = \sum_{l=1}^{m} y_l$ and, for i such that $\sum_{l=1}^{r-1} y_l < i \leq \sum_{l=1}^{r} y_l$ (note that for a given r there will be y_r indices i satisfying this condition), let A_i and B_i be the sets A and B (from either (5) or (6)) that correspond to the (in)equality that defines the r-th row of M. Note that (7) implies

$$\sum_{l=1}^{j} P_i(A_l) - \sum_{l=1}^{j} P_i(B_l) = 0 \quad \text{for all} \quad i = 1, \ldots, k \tag{8}$$

which is equivalent to (3). Also note that by definition of A_i and B_i, we have $A_i \succeq B_i$ for all $i = 1, \ldots j$. Furthermore, because $y_1 > 0$ we have $A_1 \succ B_1$. Therefore, if such $\mathbf{y}$ existed, the cancellation condition would not hold. Thus, we conclude that such $\mathbf{y}$ does not exist, completing the sufficiency part of the proof.

Conversely, if $\succ$ can be represented by (4), then for any $A_1, \ldots, A_j$ and $B_1, \ldots, B_j$, using

(8) in the third equality, we get

$$\begin{aligned}
\sum_{l=1}^{j} u(A_l) &= \sum_{l=1}^{j}\sum_{i=1}^{k} w_i P_i(A_l) \\
&= \sum_{i=1}^{k} w_i \sum_{l=1}^{j} P_i(A_l) \\
&= \sum_{i=1}^{k} w_i \sum_{l=1}^{j} P_i(B_l) \\
&= \sum_{l=1}^{j}\sum_{i=1}^{k} w_i P_i(B_l) \\
&= \sum_{l=1}^{j} u(B_l).
\end{aligned}$$

Thus, if $A \succ B$, i.e, if $u(A_1) > u(B_1)$, there has to exists m such that $u(B_m) > u(A_m)$, i.e., such that $B_m \succ A_m$. ▮

Two special cases of the cancellation condition and the corresponding result of Theorem 5 are well-known. First, for P_x, $x \in [n]$, defined as $P_x(A) = 1$ if and only if $x \in A$, the cancellation condition is a necessary and sufficient condition for an additive utility function (introduced in Example 2) to exist: see, e.g.,[14]. The case where set properties are extended to also include two-element sets, i.e., all P_S, $|S| \leq 2$, defined as $P_S(A) = 1$ if and only if $S \subseteq A$, has been the topic of [22].

Theorem 5 also provides a characterization for the existence of a multiplicative representation u.

Corollary 6 *A preference relation $\succ$ satisfies the cancellation condition for the set proper-*

ties $P_1, \ldots, P_k$ if and only if there exist real numbers $w_i > 0$ such that

$$u = \prod_{i=1}^{k} w_i P_i \tag{9}$$

is a representation of $\succ$.

Proof. For any $A \subseteq [n]$ representation (9) can be rewritten as

$$u(A) = \prod_{i, P_i(A)=1} w_i$$

and, thus,

$$\log u(A) = \sum_{i, P_i(A)=1} \log(w_i) = \sum_{i=1}^{k} \log(w_i) P_i(A).$$

Further note that, $u(A) > u(B)$ if and only if $\log u(A) > \log u(B)$. Thus, representation (9) exists if and only if representation $u = \sum_{i=1}^{k} log(w_i) P_i$ exists, and the latter is characterized by Theorem 5. ∎

3 Comments

We have characterized the existence of a linear representation of a weak order $\succ$ on the set of all subsets of a finite set (Theorem 5). While a linear representation with at most $2^n - 1$ terms always exists, there is no guarantee that such representation is succinct (Proposition 3).

Identifying a small number of properties that determine subset valuations is critical in many practical settings; one such setting was illustrated in Example 4. In fact, the tractability and performance of data-driven models for subset valuations critically depends on the model specification. For example, regression-based machine learning approaches in big data settings require a limited number of covariates (which can be viewed as set properties), so parsimony and tractable model specification is essential, even if the set of model covariates is suboptimal (e.g., not minimal). Set properties defining subset valuations need not be unique, and yield a theoretically interesting open question: for a weak order $\succ$ on $2^{[n]}$ (possibly with some additional properties of practical interest), determine an upper bound k so that there exists a representation u which is a linear combination of at most k set property functions. Theorem 5 either verifies that such function u is indeed a representation of subset valuations, or provides a family of sets which demonstrate the misrepresentation via our Cancellation Condition.

It is plausible that functional formats more complex than linear format considered in this paper, could also yield succinct representations. However, linear functional forms are fundamental to regression-based models as data is used to specify the function by determining optimal coefficients (i.e., weights for set properties). Further, linear functionals are best local approximations for any differentiable functional form. This suggests that a generic characterization of linear representations of subset valuations presented here is a natural starting point in the analysis of possible representations utilizing any other functional forms. Additionally, analyzing linear forms allows for utilizing fundamental duality theorems. The key to defining our Cancellation Condition and the main argument in the proof of Theorem 5 are both based on one of many equivalent statements of the theorem of the alternative (i.e., the Farkas Lemma) which *tells us than any inconsistent set of "axioms" can be refuted by a suitable derivation. . . . This view makes the Farkas Lemma a (small) cousin of various completeness theorems of logic and of other famous result, such as Hilbert's Nullstellensatz* *in algebraic geometry* ([28]). So, our Cancellation Condition is a natural representation of the theorem of the alternative: either the linear representation exist or there is inconsistency manifested in the violation of the Cancellation Condition. This approach in the context of social choice has been pioneered and utilized by Peter Fishburn: [15] lays out the theoretical argument, while his numerous papers, e.g. [16, 17, 18, 19, 22, 20, 21, 7], utilize variants of the argument.

Proof of Theorem 5 on its face value requires constructing exponentially many linear constraints (5) and (6). However, establishing non-existence of a linear representation only requires identifying a family of sets that violates the Cancellation Condition. One such family is in turn provided by those sets that correspond to potentially a small number of non-zero components of the dual vector y. There might be multiple other families one could construct, taking into account contextual details of a particular setting (thereby implicitly constructing the dual vector y).

Whenever there is a need to determine subset valuations, one should try first to assess the underlying factors that determine those valuations. These factors might be values of individual objects, as well as other factors (defined here as set properties) that cause either positive or negative synergetic effects. In this note, we provide necessary and sufficient conditions for these factors to be additively or multiplicatively combined into a valuation function. If a reasonably simple valuation function exists, the set properties defining this function should be viewed as basic atoms for evaluating and comparing bundles. Thus, it could make sense to take such a "basis" of these set properties into account when designing elicitation, allocation, optimization, or pricing procedures. (For example, maybe one could design an auction where basic objects that have to be allocated and priced are such set properties.) An approach of trying to determine the main structural blocks of the valuation function is a possible alternative to assessing all valuations directly and/or estimating those valuations based on (implied) values of individual objects. For instance, P_{target} and P_{target2} are such "basis" for the advertiser in Example 4. As the advertiser collects information from the advertising campaign, they could note discrepancies between realized valuations and those determined by the representation. (For example, value from Manhattan residents might be different than the value from residents of other city boroughs.) Sets where such discrepancies are the largest would be natural to guide construction of a new set property for the linear representation. (If M is a set of digits corresponding to Manhattan zipcodes, one could simply add $P_{\geq 1}(S \cap M)$.) Such myopic approach to building structural blocks of the valuation function by sequentially adding most significant sources of synergy not yet identified, could be guided by using the linear representation from the proof of Theorem 5 as they might correspond to the m^*-th linear constraint with $m^* = \arg\max y_k$. (This approach is similar to the idea of the stepwise regression procedure.) While this myopic construction of a subset valuation function is likely suboptimal, further study of such alternative approach could be an interesting research direction.

References

[1] Aarts, S., D.B. Shmoys, and A. Coy. 2023. An Interpretable Determinantal Choice Model for Subset Selection. *arXiv*: 2302.11477.

[2] Abraham, I., S. Athey, M. Babaioff, M., and M.D. Grubb. 2020. Peaches, lemons, and cookies: Designing auction markets with dispersed information. *Games and*

Economic Behavior 124, 454-477.

[3] Bakos, Y., and E. Brynjolfsson. 1999. Bundling information goods: Pricing, profits and efficiency. *Management Science* 45, 1613-1630.

[4] Benson, A.R., R. Kumar, and A. Tomkins. 2018. A discrete choice model for subset selection. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, 37-45.

[5] Bergemann, D., and A. Bonatti. 2015. Selling cookies. *American Economic Journal: Microeconomics* 7, 259-294.

[6] Bernstein, F., S. Modaresi, S., and D. Saure. 2019. A dynamic clustering approach to data-driven assortment personalization. *Management Science* 65, 2095-2115.

[7] Brams, S.J., P.H. Edelman, and P.C. Fishburn. 2003. Fair division of indivisible items. *Theory and Decision* 55, 147-180.

[8] Candogan, O., and S. Pekeč. 2018. Efficient allocation and pricing of multifeatured items. *Management Science* 64, 5521-5543.

[9] Chernev, A. 2005. Feature complementarity and assortment in choice. *Journal of Consumer Research* 31, 748-759.

[10] Cohen, M.C., I. Lobel, I., and R. Paes Leme. 2020. Feature-based dynamic pricing. *Management Science* 66, 4921-4943.

[11] Conen, W., and T. Sandholm. 2001. Preference elicitation in combinatorial auctions. *IJCAI 2001 Workshop on Economic Agents, Models, and Mechanisms*, 71-80.

[12] Cramton, P., Y. Shoham, and R. Steinberg. 2006. *Combinatorial auctions.* MIT Press.

[13] Doignon, J.P., A. Pekeč, and M. Regenwetter. 2004. The repeated insertion model for rankings: Missing link between two subset choice models. *Psychometrika* 69, 33-54.

[14] Fishburn, P.C. 1970. *Utility Theory for Decision Making.* Wiley, NY.

[15] Fishburn, P.C. 1971. The theorem of the alternative in social choice theory. *Operations Research* 19, 1323-1330.

[16] Fishburn, P.C. 1972. *Mathematics of Decision Theory.* Mouton.

[17] Fishburn, P.C. 1991. Nontransitive preferences in decision theory. *Journal of Risk and Uncertainty* 4, 113-134.

[18] Fishburn, P.C. 1992. Utility as an additive set function. *Mathematics of Operations Research* 17, 910-920.

[19] Fishburn, P.C. 1992. Signed orders and power set extensions. *Journal of Economic Theory* 56, 1-19

[20] Fishburn, P.C. 1997. Cancellation conditions for multiattribute preferences on finite sets. *Essays In Decision Making: A Volume in Honour of Stanley Zionts*, 157-167.

[21] Fishburn, P.C. 2001. Cancellation conditions for finite two-dimensional additive measurement. *Journal of Mathematical Psychology* 45, 2-26.

[22] Fishburn, P.C., and I.H. LaValle. 1996. Binary interactions and subset choice. *European Journal of Operational Research* 92, 182-192.

[23] Fishburn, P.C., A. Pekeč, and J.A. Reeds. 2002. Subset comparisons for additive linear orders. *Mathematics of Operations Research* 27, 227-243.

[24] Gallego, G., and H. Topaloglu. 2019. *Revenue Management and Pricing Analytics.* Springer.

[25] Hanson, W.A., and R.K. Martin. 1990. Optimal bundle pricing. *Management Science* 36, 155-174.

[26] Kök, A.G., M. L. Fisher, and R. Vaidyanathan. 2015. Assortment planning: Review of literature and industry practice. *Retail supply chain management: Quantitative models and empirical studies*, 175-236.

[27] Luce, R.D. 2000. *Utility of Gains and Losses: Measurement-Theoretic and Experimental Approaches*, Mahwah, NJ: Lawrence Erlbaum Association.

[28] Matoušek, J., and B. Gärtner. 2007. *Understanding and using linear programming.* Springer.

[29] Nisan, N., and I. Segal. 2006. The communication complexity of efficient allocations and supporting prices. *Journal of Economic Theory* 129, 192-224.

[30] Parkes, D.C. 2005. Auction design with costly preference elicitation. *Annals of Mathematics and Artificial Intelligence.* 44, 1012-2443.

[31] Pekeč, A., and M.H. Rothkopf. 2003. Combinatorial auction design. *Management Science* 49, 1485-1503.

[32] Rothkopf, M.H., A. Pekeč, and R.M. Harstad. 1998. Computationally manageable combinational auctions. *Management Science* 44, 1131-1147.

[33] Sandhom, T., and C. Boutilier. 2006. Preference elicitation in combinatorial auctions. Chapter 10 in Cramton, Shoham, Steinberg (Eds) *Combinatorial Auctions.* M.I.T. Press.

[34] Schrijver, A. 1998. *Theory of linear and integer programming.* Wiley.

[35] van Rooij, I, U. Stege, and H. Kadlec. 2005. Sources of complexity in subset choice. *Journal of Mathematical Psychology* 49, 160-187.

[36] Venkatesh, R., and W. A. Kamakura. 2003. Optimal bundling and pricing under a monopoly: Contrasting complements and substitutes from independently valued products. *Journal of Business* 76, 211-231.