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

Peter Fishburn* Aleksandar Saša Pekeč†

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 bundle valuations might point to simple and concise representation of the valuation function. We characterize additive and multiplicative representations of synergies in bundle valuations and subset utility, which in turn points to necessary and sufficient conditions for a succinct representation of bundle valuations to exist.

*Retired from Information Sciences Research Center, AT&T Shannon Laboratory, 180 Park Avenue, Florham Park, NJ 07932, USA.
†The Fuqua School of Business, Duke University, Durham, NC 27708-0120, USA. E-mail: pekec@duke.edu
1 Introduction

An issue that arises naturally in the context of multiple objects is that of bundling. Decisions often have to be made on whether to bundle some objects together or possibly unbundle available bundles. For example, the issue of how to create and price bundles optimally is an important problem in marketing and management science (e.g., see [7, 1, 15]). Another setting in which bundles play a central role is combinatorial auctions. In combinatorial auctions, multiple goods are auctioned simultaneously and bidders can submit bids on bundles of objects (instead of having to submit separate bids on single objects). The potentially difficult determination of combinatorial auction winners could be alleviated if bidding is restricted only on carefully chosen bundles [12].

Understanding preferences over bundles is a prerequisite for making meaningful decisions in creating, choosing, allocating or pricing bundles. (Even the problem of choosing a subset of a finite set is known to be computationally unmanageable, [14].) For example, in a combinatorial auction every bidder needs to understand its own preferences over 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 [11] that has recently stimulated a considerable research interest (e.g., [2, 9, 10, 13]).

The fundamental problem of eliciting preferences over the set of all bundles is that the number of bundles could be prohibitively large. For example, more than a million possible bundles can be created from 20 objects \(2^{20} - 1 = 1048575\) nonempty bundles, to be exact).
Hence, there is interest in methods that avoid having to elicit valuations for all possible bundles.

For example, one of the main ideas in the design and use of iterative combinatorial auction procedures is that an auction procedure only needs to elicit information about bidders’ valuations that are relevant in the context of reported demand. Another approach attempts 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 [12]), in which case bundle valuations might be represented in a succinct and simple way.

In this note we provide a necessary and sufficient condition for bundle valuations (and subset utility) to be representable as a sum (or a product) of different synergetic effects. More precisely, given a collection of properties (that each bundle could have or not have), we characterize the existence of weights, one for each such property, such that the utility of any bundle is the sum of the weights corresponding to the properties that bundle possesses. One special case, where only sources of synergy are binary interactions (i.e., synergetic effects are limited to pairs of objects that a bundle contains) has been the topic of [5]. The results presented here generalize those in [5]. However, our main point is not that results from
can be generalized, but it is that our generalization indicates that correctly identifying sources of synergy could yield to simple bundle 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 bundles 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 [3], \(\succ\) 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 $u : 2^{[n]} \to \mathbb{R}$, $u(\emptyset) = 0$, such that

$$A \succ B \iff u(A) > u(B).$$

(1)

The fact that a finite weak order can have a representation in real numbers is well-known, e.g., see [3].

Note that the representation $u$ from Proposition 1 is not unique: for any $f : \mathbb{R} \to \mathbb{R}$ 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 $x \in [n]$.

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 \( P_{\text{conn}}(A) = 1 \) 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 \( S \subseteq [n] \), 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 \( P_x = 1 \) 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:

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

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 relation \( \succ \) on \( 2^{[n]} \) is a linear combination of at most \( 2^n - 1 \) set property functions. If \( \succ \) is complete, then there exist \( u \) such that \( u \) cannot be decomposed into a linear combination of less than \( 2^n - 1 \) set property functions.

**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 most representations $u$ of a complete $\succ$ cannot be decomposed into a linear combination of less than $2^n - 1$ set property functions. 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 $A_1, A_2, \ldots, A_{2^n - 1}$, we have

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

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

$$u = (u(A_1), \ldots, u(A_{2^n-1}))^T \quad \text{and} \quad w = (w_1, \ldots, w_k)^T,$$

(2) becomes

$$u = M \cdot w.$$ 

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 $\mathbb{R}^{2^n-1}$, 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 $\mathbb{R}^{2^n-1}$. 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. \[\square\]

To characterize $\succ$ that allow utility representations that 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 $A_1, A_2, \ldots, A_j$ and $B_1, B_2, \ldots, B_j$ satisfying

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

$A_1 \succ B_1$ 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 4** 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

$$u = \sum_{i=1}^{k} w_i P_i \quad (4)$$

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 $m = 2^n - 1$ 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

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

and

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

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 $\mathbf{M}$ are either -1, 0, or 1. The theorem of the alternative (or Farkas Lemma; e.g., see [4]) states that either a solution $\mathbf{w}$ exists or there exists a non-negative integral solution of

$$\mathbf{yM} = \mathbf{0} \quad (7)$$

with $y_{i^*} > 0$ for at least one $i^*$ such that the $i^*$-th row of $\mathbf{M}$ 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. Suppose that $\mathbf{y} = (y_1, \ldots, y_m)$ is 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 $\sum_{l=1}^{i-1} y_l < i \leq \sum_{l=1}^r y_l$, 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

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

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 $y$ existed, the cancellation condition would not hold. Thus, we conclude that such $y$ 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

$$
\sum_{i=1}^{j} u(A_i) = \sum_{l=1}^{j} \sum_{i=1}^{k} w_i P_i(A_l)
$$

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 $y$ existed, the cancellation condition would not hold. Thus, we conclude that such $y$ 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

$$
\sum_{i=1}^{j} u(A_i) = \sum_{l=1}^{j} \sum_{i=1}^{k} w_i P_i(A_l)
$$

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 $y$ existed, the cancellation condition would not hold. Thus, we conclude that such $y$ 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

$$
\sum_{i=1}^{j} u(A_i) = \sum_{l=1}^{j} \sum_{i=1}^{k} w_i P_i(A_l)
$$

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 $y$ existed, the cancellation condition would not hold. Thus, we conclude that such $y$ 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

$$
\sum_{i=1}^{j} u(A_i) = \sum_{l=1}^{j} \sum_{i=1}^{k} w_i P_i(A_l)
$$

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 $y$ existed, the cancellation condition would not hold. Thus, we conclude that such $y$ 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

$$
\sum_{i=1}^{j} u(A_i) = \sum_{l=1}^{j} \sum_{i=1}^{k} w_i P_i(A_l)
$$

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 $y$ existed, the cancellation condition would not hold. Thus, we conclude that such $y$ 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

$$
\sum_{i=1}^{j} u(A_i) = \sum_{l=1}^{j} \sum_{i=1}^{k} w_i P_i(A_l)
$$

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 $y$ existed, the cancellation condition would not hold. Thus, we conclude that such $y$ does not exist, completing the sufficiency part of the proof.
(introduced in Example 2) to exist: see, e.g.,[3]. 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 [5].

Theorem 4 also provides a characterization for the existence of a multiplicative representation $u$.

**Corollary 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 $w_i > 0$ such that

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

(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,

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

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 4. 

11
3 Comments

We have characterized the existence of a linear representation of a preference relation $\succ$ on the set of all subsets of a finite set (Theorem 4). While a linear representation with at most $2^n - 1$ terms always exists, there is no guarantee that such representation is succinct (Proposition 3).

Even if subset valuations can be represented in a succinct way, that does not mean that such valuations could be efficiently elicited. Further, the choice of a query type and design of the elicitation procedure is also important: optimal queries in a one-shot elicitation scheme (where a predetermined list of queries is answered simultaneously) could be suboptimal in a $k$-stage or iterative elicitation procedure (where the choice of queries in later stages could depend on the answers collected up to that point) and vice-versa. For example, as shown in [6], even if the elicitor knows that $\succ$ can be represented by an additive subset utility function $u(S) = \sum w_x p_x(S) = \sum_{x \in S} w_x$ and thus only needs to determine $n$ parameters $w_x$, and if the elicitation queries are subset relative rank ordering questions of the type "Is $u(S) > u(T)$?", an exhaustive search (which implies exponentially many queries) is required to determine $\succ$ in a one-shot approach, while there are modest shortcuts in multistage elicitation procedures.

Whenever there is a need to determine valuations of bundles, one should try first to assess the underlying factors that determine those valuations. These factors might be values of individual items in the bundle, as well as other factors (defined here as set properties) that cause either positive or negative synergetic effects. In this note we provide a necessary and a sufficient condition 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 items. The further study of such alternative approach could be an interesting research direction.

References


