OPERATIONS RESEARCH
Vol. 26, No. 2, March-April 1978

0030-364X/78/2602-0305 $01.25

Minimizing a Submodular Function on a Lattice

DONALD M. TOPKIS

Monona, Wisconsin

(Received December 1975; accepted July 1977)

This paper gives general conditions under which a collection of optimization problems, with the objective function and the constraint set depending on a parameter, has optimal solutions that are an isotone function of the parameter. Relating to this, we present a theory that explores and elaborates on the problem of minimizing a submodular function on a lattice.

WE GIVE general conditions under which a collection of optimization problems, with the objective function and the constraint set depending on a parameter, has optimal solutions that are an isotone function of the parameter. Relating to this, we present a theory that explores and elaborates on the problem of minimizing a submodular function on a lattice.

Formally stated, the main question involves the collection of optimization problems

$$\min f(x, t), \quad x \in S_t \tag{1}$$

where the variable is x and both the constraint set S_t and the objective function $f(x, t)$ depend upon the parameter t, with t being a member of the parameter set T. If S_t^* is the set of optimal solutions for (1) given t in T, then the object is to find conditions under which one can select s_t in S_t^* for each t in T such that s_t is an isotone function of t, i.e., $t \leqq b$ in T implies $s_t \leqq s_b$.

One application considers the problem of finding minimum cost paths from one node to all other nodes in an acyclic network. The network has nodes $1, \cdots, N$, an arc may only go from node i to node j if $i<j$, and the cost associated with such an arc is $c(i, j)$. Let $f(j)$ be the cost of a minimum cost path from node 1 to node j; hence $f(j)$ may be found from the well-known recursion

$$f(j) = \min_{1 \leqq i \leqq j-1} (f(i) + c(i, j)), \qquad f(1) = 0. \tag{2}$$

The recursion (2) also occurs in finding optimal solutions for several dynamic economic lot-size models in inventory theory, as in the papers of Eppen, Gould and Pashigian [6], Wagner and Whitin [26], and Zangwill

[27]. If $i(j)$ is an optimal i given j in (2), then one would want to know when $i(j)$ is isotone in j.

Section 1 introduces some standard concepts involving the notions of order and lattices. Ordering is of course fundamental throughout and it is used for comparing elements of T, for comparing selections s_t, for considering the existence of an isotone selection, in selecting an element from S_t, and in defining certain properties of the objective function $f(x, t)$. The constraint sets S_t will subsequently be assumed to be sublattices, and the sets of optimal solutions S_t^* will turn out to be sublattices.

Section 2 provides some properties of a natural partial ordering relation on the collection of nonempty sublattices of a given lattice. This relation orders the constraint sets S_t and the sets of optimal solutions S_t^*. If the sets S_t^* are isotone in t with respect to this order, then Theorem 2.3 indicates how to select s_t from S_t^* so that s_t is isotone in t.

To establish the existence of isotone optimal solutions for (1), we will assume that the objective function $f(x, t)$ has properties of sumodularity and antitone differences. Section 3 explores these properties in detail. A submodular function on a product set is characterized in terms of antitone differences, with the latter often being an easier property to recognize and conceptualize. Antitone differences are closely related to the economic concept of complementary products. Several common operations generate or preserve submodularity.

Section 4 gives conditions for each set of optimal solutions S_t^* to be a sublattice and for each to have a greatest and a least element. The minimization operation preserves submodularity and this gives conditions for the optimal value of the objective function in (1) to be a submodular function of t.

Section 5 shows that a submodular function on a sublattice of a lattice can often be extended to a submodular function on the entire lattice. In terms of (1) this can mean that there may be no loss of generality in considering the objective function $f(x, t)$ to have its essential properties on $(\bigcup_{t \in T} S_t) \times T$ rather than the more limited domain $\bigcup_{t \in T}(S_t, t)$. Thus if $c(i, j)$ is submodular on the acyclic network, then it is possible to consider a submodular cost function defined on all arcs of the full network that takes on the same values $c(i, j)$ on those arcs of the original acyclic network. This result may be convenient in applying the main isotonicity results of Theorems 6.1 and 6.2.

Section 6 gives the general isotonicity results, which provide conditions for the optimal solution sets S_t^* to be isotone in t and for the existence of optimal solutions s_t in S_t^* such that s_t is isotone in t. One application of this result is that in the acyclic network example $i(j)$ will be isotone in j if $c(i, j)$ is submodular, and this yields several known results from inventory theory [6, 26, 27] as special cases. Section 7 notes some analogies between submodular functions and convex functions.

A series of papers applies this theory to graph theory and network flows [21], n-person games [20], dynamic stochastic decision theory, inventory theory, and other structured mathematical programming problems [23]. These papers consider specific applied questions whose answers follow directly from this theory. For example, how are solutions to optimization problems in a network affected as the graph structure is modified? In an n-person, nonzero-sum game, how does the optimal strategy for one player depend on the other players' strategies and when does such dependence imply the existence of an equilibrium point and provide an algorithm for finding one? In certain structured dynamic stochastic decision problems, how will the optimal decision in each time period depend on the state of the system at that time? In various deterministic and stochastic inventory problems, how do the optimal policies depend on the costs, the demands, the bounds on the variables, and the initial inventory level? Veinott (personal communication) has also found applications of this theory. A variety of existing results established by more specialized methods also follow from our results. We will give several brief illustrative applications in the context of specific results.

1. DEFINITIONS, NOTATION, AND RELATED BACKGROUND

A *partially ordered set* (*poset*) is a set on which there is a binary relation $\leqq$ that is reflexive, antisymmetric and transitive. If $x \leqq y$ and $x \neq y$, then $x<y$ is written. Two elements x and y of a poset are *unordered* if neither $x \leqq y$ nor $y \leqq x$. A poset is a *chain* if it does not contain an unordered pair of elements. An element x of a poset S is the *greatest* (*least*) element of S if $y \leqq x (x \leqq y)$ for all y in S. The *power set*, $P(S)$, is the set of all nonempty subsets of a poset S.

If two elements, x and y, of a poset have a least upper bound (greatest lower bound), denoted $x \vee y (x \wedge y)$, it is their *join* (*meet*). A poset that contains the join and the meet of each pair of its elements is a *lattice*. If T is a subset of a lattice S and T contains the join and meet (with respect to S) of each pair of elements of T, then T is a *sublattice* of S. If a nonempty subset T of a poset S has a least upper bound (greatest lower bound) in S, denoted sup T (inf T), then this element is the *supremum* (*infimum*) of T. A lattice in which every nonempty subset has a supremum and infimum is *complete*. The set of all nonempty sublattices of a lattice S is denoted $L(S)$.

If S_α is a poset with relation $\leqq_\alpha$ for each $\alpha \in A$, then the *direct product* of these posets is the poset consisting of the set $\times_{\alpha \in A} S_\alpha = \{x = (x_\alpha) : x_\alpha \in S_\alpha$ for all $\alpha \in A\}$ with the relation $\leqq$ where $x \leqq y$ if $x_\alpha \leqq_\alpha y_\alpha$ for each $\alpha \in A$. The direct product of lattices is a lattice. If S is a subset of $X \times T$, then the *section* of S at $t \in T$ is $S_t = \{x : (x, t) \in S\}$ and the *projection* of S on T is $\Pi_T S = \{t : S_t$ is nonempty$\}$. If X and T are lattices and S is a sublattice of $X \times T$, then each section S_t is a sublattice of X and the projection $\Pi_T S$

is a sublattice of T. Bergman [2] and Topkis [24] have explored and characterized the structure of sublattices of the product of n lattices; see also Baker and Pixley [1]. In [24] the results on sublattice structure lead to analogous results characterizing the structure of those functions whose level sets are sublattices.

A function f from a poset S to a poset T is *isotone* (*antitone*) if $x \leqq y$ in S implies $f(x) \leqq f(y)$ $(f(y) \leqq f(x))$ in T. A function f from a poset S to a poset T is *strictly isotone* (*strictly antitone*) if $x < y$ in S implies $f(x) < f(y)$ $(f(y) < f(x))$ in T.

2. A PARTIAL ORDERING ON $L(S)$

Suppose a lattice S with relation $\leqq$ is given. For X and Y in the power set $P(S)$, X is *lower than* Y, written $X \leqq^p Y$, if $x \in X$ and $y \in Y$ imply that $x \wedge y \in X$ and $x \vee y \in Y$. Veinott (personal communication) introduced this relation.

LEMMA 2.1. *For a lattice S, the relation $\leqq^p$ is antisymmetric and transitive on $P(S)$.*

Proof. Pick any X, $Y \in P(S)$ for which $X \leqq^p Y$ and $Y \leqq^p X$. Now pick any $x \in X$ and $y \in Y$. Because $X \leqq^p Y$, $x \wedge y \in X$ and $x \vee y \in Y$. But then because $Y \leqq^p X$, $y = y \vee (x \wedge y) \in X$ and $x = (x \vee y) \wedge x \in Y$. Thus $X = Y$ and $\leqq^p$ is antisymmetric on $P(S)$.

Pick any X, Y, $Z \in P(S)$ such that $X \leqq^p Y$ and $Y \leqq^p Z$. Now pick any $x \in X$, $y \in Y$, and $z \in Z$. Since $X \leqq^p Y$ and $Y \leqq^p Z$, $x \vee y \in Y$ and $y \wedge z \in Y$. Thus $x \vee (y \wedge z) \in Y$ and so $x \vee z = x \vee ((y \wedge z) \vee z) = (x \vee (y \wedge z)) \vee z \in Z$ because $Y \leqq^p Z$. Similarly, $(x \vee y) \wedge z \in Y$ and so $x \wedge z = (x \wedge (x \vee y)) \wedge z = x \wedge ((x \vee y) \wedge z) \in X$ because $X \leqq^p Y$. Therefore, $X \leqq^p Z$ and $\leqq^p$ is transitive on $P(S)$.

For a lattice S it follows directly from the definition of a sublattice that $X \in L(S)$ if and only if $X \leqq^p X$. This observation together with Lemma 2.1 yields the following result.

THEOREM 2.1. *If S is a lattice with the relation $\leqq$, then $L(S)$ is a poset with the relation $\leqq^p$.*

If S is a lattice, T is a poset and S_t is an isotone (antitone) function from T into the poset $L(S)$ with relation $\leqq^p$ (hence $t \leqq b$ in T implies $S_t \leqq^p S_b (S_b \leqq^p S_t)$ in $L(S)$), then S_t is *ascending* (*descending*) in t on T.

If S is a lattice, then $\{x : x \in S,\ x \leqq t\}$ and $\{x : x \in S,\ t \leqq x\}$ are ascending in t on S.

THEOREM 2.2. *If S_{zt} is ascending in t on T for each z in an arbitrary set Z and if $\bigcap_{z \in Z} S_{zt}$ is nonempty for each t in T, then $\bigcap_{z \in Z} S_{zt}$ is ascending in t on T.*

Proof. Pick any b and t in T with $t \leqq b$. Pick any $x \in \bigcap_{z \in Z} S_{zt}$ and $y \in \bigcap_{z \in Z} S_{zb}$. Then $x \in S_{zt}$ and $y \in S_{zb}$ for each $z \in Z$ and so by hypothesis $x \wedge y \in S_{zt}$ and $x \vee y \in S_{zb}$ for each $z \in Z$. Thus $x \wedge y \in \bigcap_{z \in Z} S_{zt}$ and $x \vee y \in \bigcap_{z \in Z} S_{zb}$.

Let S be a lattice and X a nonempty sublattice of S. By applying Theorem 2.2 together with the preceding example and the earlier observation that a nonempty sublattice is lower than itself, we see that $S_{rt} = \{x : x \in X,\ r \leqq x,\ x \leqq t\}$ is ascending in (r, t) on the sublattice $\{(r, t) : S_{rt}$ is nonempty$\}$ of $S \times S$.

THEOREM 2.3. *If S is a complete lattice, X and Y are in $P(S)$, and $X \leqq^p Y$, then* $\inf X \leqq \inf Y$ and $\sup X \leqq \sup Y$.

Proof. For any $x \in X$ and $y \in Y$, $x \wedge y \in X$ and $x \vee y \in Y$ because $X \leqq^p Y$. Therefore, $\inf X \leqq x \wedge y \leqq y$ and $x \leqq x \vee y \leqq \sup Y$, and hence $\inf X \leqq \inf Y$ and $\sup X \leqq \sup Y$.

3. SUBMODULAR FUNCTIONS ON A LATTICE

Suppose f is a real-valued function on a lattice S. If

$$f(x \wedge y) + f(x \vee y) \leqq f(x) + f(y) \tag{3}$$

for all x and y in S, then f is *submodular* on S. If $f(x \wedge y) + f(x \vee y) < f(x) + f(y)$ for all unordered x and y in S, then f is *strictly submodular* on S. If $-f$ is (strictly) submodular, then f is (*strictly*) *supermodular.*

Suppose X and T are posets and f is a real-valued function on $S \subseteq X \times T$. If $f(x, z) - f(x, t)$ is isotone, antitone, strictly isotone, or strictly antitone in x on $S_z \cap S_t$ for each $t < z$ in T, then f has, respectively, *isotone differences, antitone differences, strictly isotone differences,* or *strictly antitone differences* in (x, t) on S. The conditions of these definitions do not distinguish between the first and second variables because $f(x, z) - f(x, t) \leqq f(y, z) - f(y, t)$ if and only if $f(y, t) - f(x, t) \leqq f(y, z) - f(x, z)$ — and similarly for a strict inequality.

Suppose $S_1, \cdots, S_n$ are posets, $S \subseteq \times_{i=1}^n S_i$, an element x in S is expressed as $x = (x_1, \cdots, x_n)$ where $x_i \in S_i$ for each i, and f is a real-valued function on S. If, on S, f has isotone differences, antitone differences, strictly isotone differences, or strictly antitone differences in (x_j, x_k) for all $j \neq k$ with each x_i fixed for $i \neq j$ and $i \neq k$, then f has, respectively, *isotone differences, antitone differences, strictly isotone differences,* or *strictly antitone differences* on S.

Multiplication by a positive scalar and the addition of functions preserve each of the properties defined above.

THEOREM 3.1. If *S_i is a lattice for $i = 1, \cdots, n$, S is a sublattice of $\times_{i=1}^n S_i$, and f is* (*strictly*) *submodular on S, then f has* (*strictly*) *antitone differences on S.*

The proof of Theorem 3.1 is immediate from the definitions. Since the proof of Theorem 3.2 for the case of strictly antitone differences indicated in parentheses is only slightly different from the case of antitone differences, the proof of the former is omitted.

THEOREM 3.2. *If* S_i *is a chain for* $i=1, \cdots, n$ *and* f *has (strictly) antitone differences on* $\times_{i=1}^{n} S_i$, *then* f *is (strictly) submodular on* $\times_{i=1}^{n} S_i$.

Proof. Pick any x and y in $\times_{i=1}^{n} S_i$. If $x \leqq y$ or $y \leqq x$, then (3) holds as an equality; so suppose x and y are unordered. For convenience arrange the components of x and y so that $x \wedge y = (x_1, \cdots, x_k, y_{k+1}, \cdots, y_n)$ and $x \vee y = (y_1, \cdots, y_k, x_{k+1}, \cdots, x_n)$ for some k, $0 \leqq k \leqq n$. Such an arrangement is possible because each S_i is a chain. Because x and y are unordered, $0<k<n$. For $0 \leqq i \leqq j \leqq n$ define $z^{ij} = (y_1, \cdots, y_i, x_{i+1}, \cdots, x_j, y_{j+1}, \cdots, y_n)$. If $0 \leqq i < k \leqq j < n$, then $x_{i+1} \leqq y_{i+1}$ in S_{i+1} and $y_{j+1} \leqq x_{j+1}$ in S_{j+1}; hence the property of antitone differences implies $f(z^{i+1,j+1}) - f(z^{i,j+1}) \leqq f(z^{i+1,j}) - f(z^{i,j})$. Thus for $k \leqq j < n$

$$\begin{aligned} f(z^{k,j+1}) - f(z^{0,j+1}) &= \sum_{i=0}^{k-1} \left(f(z^{i+1,j+1}) - f(z^{i,j+1})\right) \\ &\leqq \sum_{i=0}^{k-1} \left(f(z^{i+1,j}) - f(z^{i,j})\right) = f(z^{k,j}) - f(z^{0,j}). \end{aligned}$$

Therefore, $f(x \vee y) - f(x) = f(z^{k,n}) - f(z^{0,n}) \leqq f(z^{k,k}) - f(z^{0,k}) = f(y) - f(x \wedge y)$, which is the desired inequality.

The result of Theorem 3.2 is not valid for the product of a countable collection of chains. Let $S_i = \{0, 1\}$ where $0<1$ for $i=1, 2, \cdots$. Define $f(x)$ on $\times_{i=1}^{\infty} S_i$ so that $f(x)=0$ if $x_i=1$ for an infinite set of indices i and $f(x)=1$ if $x_i=1$ for only a finite set of indices. With respect to any finite set of variables $f(x)$ is constant and so has both antitone differences and isotone differences. However, $f(x)$ is not submodular on $\times_{i=1}^{\infty} S_i$ since if $\bar{x}$ and $\bar{y}$ are defined by $\bar{x}_i=1$ if i is odd and $\bar{x}_i=0$ if i is even, $\bar{y}_i=1$ if i is even and $\bar{y}_i=0$ if i is odd, then $f(\bar{x} \wedge \bar{y}) + f(\bar{x} \vee \bar{y}) = 1 > 0 = f(\bar{x}) + f(\bar{y})$.

Theorems 3.1 and 3.2 reduce the question of submodularity on a finite product of chains to that of exploring all pairs of cross-differences. Let u^i be the ith unit vector in E^n. A function f is submodular on E^n if and only if $f(x+\epsilon u^i) - f(x)$ is antitone in x_j for all $i \neq j$, $\epsilon > 0$, and x. If f is differentiable on E^n, then f is submodular if and only if $\partial f(x)/\partial x_i$ is antitone in x_j for all $i \neq j$ and x. If f is twice differentiable on E^n, then f is submodular if and only if $\partial^2 f(x)/\partial x_i \partial x_j \leqq 0$ for all $i \neq j$ and x. Antitone differences—and equivalently submodularity—are a well-known condition for a cost function to be that of a system of complementary products, as noted by Samuelson [15]. Suppose that f is the cost function (or minus the utility function) for a system of n products whose levels are $x=(x_1, \cdots, x_n)$. Then $f(x+\epsilon u^i) - f(x)$ is the additional cost for an additional $\epsilon > 0$ units. Antitone differences for f are equivalent to the property that the net additional cost for additional product i will not increase if there is more of product j, $j \neq i$—that is, the desirability of more product i will never decrease if there is an increase in product j. This is certainly a reasonable condition for complementarity. Samuelson [15] criticized this because a strictly isotone transformation may not preserve the property of antitone differences. However, the isotonicity results of Section 6 show that this property implies another property, intimately related to the notion of complementarity, which a strictly isotone transformation preserves. The relationship between submodularity and complementary products provides economic interpretations of the statements of Theorems 4.3 and 6.1. An application in Section 6 to price dependency further develops this relationship.

It is usually more natural to approach and understand applications in terms of antitone differences and complementarity, while the attendant mathematical analyses tend to be easier to handle in terms of submodularity. One problem in which submodularity arises directly is that of finding a minimum cut in a capacitated network [8]. The late Professor O. Ore pointed out (personal communication) that this problem is one of minimizing a submodular function on a lattice.

A real-valued function f on $\times_{i=1}^n S_i$ is *separable* if $f(x) = \sum_{i=1}^n f_i(x_i)$ for all $x = (x_1, \cdots, x_n)$ with $x_i \epsilon S_i$ for $i = 1, \cdots, n$. A function that is both submodular and supermodular is a *valuation*.

THEOREM 3.3. *If S_i is a chain for $i=1, \cdots, n$, then f is separable on $\times_{i=1}^n S_i$ if and only if f is a valuation on $\times_{i=1}^n S_i$.*

Proof. Suppose f is separable and x and y are in $\times_{i=1}^n S_i$. Then since S_i is a chain, $x_i \wedge y_i = x_i$ and $x_i \vee y_i = y_i$ or the reverse is true. Thus $f(x \wedge y) + f(x \vee y) = \sum_{i=1}^n (f_i(x_i \wedge y_i) + f_i(x_i \vee y_i)) = \sum_{i=1}^n (f_i(x_i) + f_i(y_i)) = f(x) + f(y)$; hence f is a valuation on $\times_{i=1}^n S_i$.

Now suppose f is a valuation on $\times_{i=1}^n S_i$. By Theorem 3.1 f has both isotone and antitone differences. Pick some fixed z in $\times_{i=1}^n S_i$. Now choose any x in $\times_{i=1}^n S_i$. Then

$$\begin{aligned} f(x) &= f(z) + \textstyle\sum_{i=1}^n (f(x_1, \cdots, x_i, z_{i+1}, \cdots, z_n) - f(x_1, \cdots, x_{i-1}, z_i, \cdots, z_n)) \\ &= f(z) + \textstyle\sum_{i=1}^n (f(z_1, \cdots, z_{i-1}, x_i, z_{i+1}, \cdots, z_n) - f(z)), \end{aligned}$$

which is separable.

The following gives a procedure for generating members of a class of submodular functions from others of that class.

LEMMA 3.1. *If S is a lattice and $g_i(x)$ is nonpositive, isotone (antitone), and submodular on S for $i=1, \cdots, k$, then $f(x) = (-1)^{k-1} g_1(x) g_2(x) \cdots g_k(x)$ is also nonpositive, isotone (antitone), and submodular on S.*

Proof. That f is nonpositive and isotone (antitone) is immediate. For $k=2$ and any x and y in S,

$$\begin{aligned} f(x\vee y)-f(x) &= -g_1(x\vee y)g_2(x\vee y)+g_1(x)g_2(x)\\ &\leqq -(g_1(x)+g_1(y)-g_1(x\wedge y))g_2(x\vee y)+g_1(x)g_2(x)\\ &\leqq -g_1(y)g_2(x\vee y)+g_1(x\wedge y)g_2(x)\\ &\leqq -g_1(y)(g_2(x)+g_2(y)-g_2(x\wedge y))+g_1(x\wedge y)g_2(x)\\ &\leqq -g_1(y)g_2(y)+g_1(x\wedge y)g_2(x\wedge y)\\ &= f(y)-f(x\wedge y). \end{aligned}$$

If we let f_k indicate the dependence of f on k, the result follows by induction for $k>2$ by observing that $f_k(x)=-f_{k-1}(x)g_k(x)$.

TABLE I

WAYS OF CONSTRUCTING A SUBMODULAR OR SUPERMODULAR FUNCTION

f convex	f concave	g supermodular	g submodular	f antitone	f isotone	g isotone (antitone)	$f\circ g$ supermodular	$f\circ g$ submodular
X		X			X	X	X	
X			X	X		X	X	
	X	X		X		X		X
	X		X		X	X		X

Table I gives a variety of ways of constructing a submodular or supermodular function by taking the composition of two other functions with certain properties. In this table one should read across any given row, and the assumptions corresponding to the checked columns on the left of the double line imply the result corresponding to the checked column to the right of the double line. All properties of g and $f\circ g$ hold on a given lattice S, g is a real-valued function on S, and f is a real-valued function on the real line. Each result is a consequence of the following identity for all x and y in S:

$$\begin{aligned} &f(g(x))+f(g(y))-f(g(x\vee y))-f(g(x\wedge y))\\ &\quad=(f(g(x))-f(g(x\vee y)+g(x\wedge y)-g(y)))\\ &\quad+(f(g(x\vee y)+g(x\wedge y)-g(y))-f(g(x\vee y))-f(g(x\wedge y))\\ &\quad+f(g(y))). \end{aligned}$$

4. MINIMIZING A SUBMODULAR FUNCTION

The following result shows that the set of points at which a subadditive function attains its minimum is a sublattice. Ore [12, p. 110] has given a completely general proof of this result in the particular context of finding a maximal deficiency in a network. Ford and Fulkerson [8] and Shapley [16] have shown that the union and the intersection of two minimum cuts (defined in terms of nodes rather than arcs) in a capacitated network are also minimum cuts, and Ore pointed out (personal communication) that this result is a direct consequence of Theorem 4.1.

THEOREM 4.1. *If f is submodular on a lattice S, then the set S^* of points at which f attains its minimum on S is a sublattice of S.*

Proof. Pick any x and y in S^*. Because f is submodular on the lattice S and $x, y \in S^* \subseteq S$, $0 \leqq f(x \vee y) - f(x) \leqq f(y) - f(x \wedge y) \leqq 0$.

In considering subsequent optimization problems, it will be useful to have a consistent method of selecting one element from the set of optimal solutions. One way is to pick that set's greatest or least element if these are known to exist. Corollary 4.1 provides conditions for the set of optimal solutions to have a greatest and a least element. Theorem 6.2 will apply that result to select a particular optimal solution corresponding to each possible parameter value in a parameterized collection of optimization problems. In this regard, the following topological conditions are introduced. The most useful case is that a sublattice of E^n must be complete if it is compact or complete in the usual topology.

In a poset S define $[x, \infty) = \{y : y \in S, x \leqq y\}$ and $(-\infty, x] = \{y : y \in S, y \leqq x\}$ for $x \in S$. Frink [9] defined the *interval topology* on a poset S as the topology for which the sets $[x, \infty)$ and $(-\infty, x]$ for $x \in S$ together with itself form a sub-basis of the closed sets. He showed that a complete lattice is compact in the interval topology. Conversely, Birkhoff [4] showed that a lattice compact in the interval topology is complete. Modifying an observation of Frink [9], Birkhoff [3] states that the interval topology on $\times_{\alpha \in A} S_\alpha$ is equivalent to the product topology (the topology of pointwise convergence) if each S_α is a lattice with a greatest and a least element and with the interval topology. If two topologies defined on a given set are such that each closed set in one topology is also closed in the other, then the latter topology is *finer*. The product topology is finer than the interval topology, but they may not be identical if each S_α does not have a greatest and a least element. For example, in $E^2 = E^1 \times E^1$ with the usual order relation, the set $\{y : y_1 \geqq 0\}$ is closed in the product topology but not in the interval topology. In [25], I have also explored certain relationships between completeness and topological properties, and I have given a modified notion of completeness that is equivalent to compactness in the product topology.

COROLLARY 4.1. *If S is a nonempty lattice that is compact in a topology finer than the interval topology and f is submodular and lower semicontinuous on S, then the set S^* of points at which f attains its minimum on S is a non-*

empty compact and complete sublattice of S and hence has a greatest and a least element.

Proof. By Theorem 4.1, S^* is a sublattice of S. Since f is lower semicontinuous and S is nonempty and compact, S^* is nonempty and compact [14]. Thus by a result of Birkhoff [4; p. 250], S^* is complete.

THEOREM 4.2. *If f is strictly submodular on a lattice S, then the set S^* of points at which f attains its minimum on S is a chain.*

Proof. Pick any x and y in S^*. Suppose x and y are unordered. Then because f is strictly submodular on the lattice S and $x, y \in S^* \subseteq S$, $0 \leqq f(x \vee y) - f(x) < f(y) - f(x \wedge y) \leqq 0$, which is a contradiction. Thus S^* has no unordered elments and so it is a chain.

THEOREM 4.3. *If X and T are lattices, S is a sublattice of $X \times T$, f is submodular on S, S_t is the section of S at $t \in T$, and $g(t) = \inf_{x \in S_t} f(x, t)$ is finite on the projection $\Pi_T S$, then $g(t)$ is submodular on $\Pi_T S$.*

Proof. Pick any t, $b \in \Pi_T S$, $x \in S_t$ and $y \in S_b$. Because S is a sublattice, $(x \vee y,\ t \vee b) = (x,\ t) \vee (y,\ b) \in S$ *and* $(x \wedge y,\ t \wedge b) = (x,\ t) \wedge (y,\ b) \in S$. *Thus*

$$g(t \vee b) + g(t \wedge b) \leqq f(x \vee y, t \vee b) + f(x \wedge y, t \wedge b) \leqq f(x, t) + f(y, b). \quad (4)$$

Now taking the infimum of the right-hand side of (4) over $x \in S_t$ and $y \in S_b$ gives the desired result.

Succinctly phrased, Theorem 4.3 states that the minimization operation preserves submodularity. When viewed in the context of the complementary product interpretation of Section 3, this is not at all surprising. Then this result simply says that after optimizing a system of complementary products with respect to certain of the products, the remaining products will still be complementary.

Erlenkotter [7] and Shapley [17] have shown that the optimal value of the objective function in the transportation problem has isotone differences with respect to the demands and minus the supplies. Their result also follows from applying Theorem 4.3 together with Theorems 3.1 and 3.2 to the dual of the transportation problem.

Theorem 4.3 is a so convenient [19, 23] in showing the preservation of submodularity from period to period in sequential decision problems. One special case of this is in Karlin's analysis of parametric variations of the stochastic demands in inventory problems [10].

5. EXTENSION OF SUBMODULAR FUNCTIONS

If f is a function on T and $T \subseteq S$, then a function g on S is an *extension* of f to S if $g(x) = f(x)$ for all x in T. In this section we will give conditions under which a submodular function on a sublattice T of a lattice S can be extended to a submodular function on S.

If S is a poset, then a subset T of S is an *upper* subset of S if $T \cap [y, \infty)$ is nonempty for each y in S. If T is a subset of a poset S and S has a greatest element z, then T is an upper subset of S if and only if z is in T. If f is a real-valued function on $T \subseteq S$, then f is *lower bounded relative to* S if f is bounded below on $T \cap [y, \infty)$ for each y in S.

LEMMA 5.1. *If S is a lattice, T is an upper sublattice of S, and f is an isotone submodular function on T that is lower bounded relative to S, then there exists an isotone submodular extension of f to S.*

Proof. For y in S, let $g(y) = \inf\{f(x) | x \in T \cap [y, \infty)\}$. Because T is an upper subset and f is lower bounded relative to S, $g(y)$ exists and is finite. Because f is isotone on T, $g(y) = f(y)$ for y in T and so g is indeed an extension. By construction, $g(y)$ is isotone on S. Define $L = \{(x, y) : x \in T, y \in S, y \leqq x\}$. Then L is a sublattice of $T \times S$ and $g(y) = \inf_{x \in L_y} f(x)$ is submodular on $S = \Pi_S L$ by Theorem 4.3.

If f is isotone and T has a least element, then clearly f is lower bounded relative to S. However, if $S = [0, 1]$, $T = (0, 1]$, and $f(x) = -1/x$ for x in T, then f is not lower bounded relative to S and there is no isotone extension to S.

If f is a real-valued function on a subset T of a lattice S, then f is *potentially isotone on T relative to S* if there exists a valuation h on S such that $f + h$ is isotone on T. If $S = T$, then f is *potentially isotone.*

A bounded function on a square in E^2 need not be potentially isotone. Define f on $[0, 1] \times [0, 1]$ so that $f(x) = 1$ if $x_1 = 0$ and x_2 is rational and $f(x) = 0$ otherwise. If f were potentially isotone, then by Theorem 3.3 there would exist real-valued functions h_1 and h_2 on $[0, 1]$ such that $g(x) = f(x) + h_1(x_1) + h_2(x_2)$ is isotone on $[0, 1] \times [0, 1]$. By fixing $x_1 = 1$, $h_2(x_2)$ must be isotone on $[0, 1]$. By fixing $x_1 = 0$, $h_2(z) - h_2(w) \geqq 1$ for $z > w$ in $[0, 1]$ with z irrational and w rational. Since these two properties are incompatible, f is not potentially isotone.

A valuation on a lattice is potentially isotone. A continuously differentiable function on a compact subset T of E^n is potentially isotone on T relative to E^n. This follows by constructing the valuation to be the linear function whose jth coefficient is minus the infimum over T of the jth partial derivative of the function.

If S is a finite sublattice of E^n, T is any subset of S, and f is any real-valued function on T, then f is potentially isotone on T relative to S. More generally, a real-valued function on a subset T of a finite modular lattice S is potentially isotone, where a lattice S is *modular* if $x \vee (y \wedge z) = (x \vee y) \wedge z$ for all $x \leqq z$ in S and $y \in S$. This follows by letting the appropriate valuation be a large enough constant times the height function on S, where the *height* of an element x in S is one less than the number of elements in the shortest chain joining x and the least element of S. The height function is a strictly isotone valuation on a finite modular lattice [4].

THEOREM 5.1. *If S is a lattice, T is an upper sublattice of S, T has a least element, and f is submodular on T and potentially isotone on T relative to S, then there exists a submodular extension of f to S.*

Proof. Because f is submodular on T and potentially isotone on T relative to S, there exists a valuation h on S such that $f+h$ is submodular and isotone on T. Because T has a least element and $f+h$ is isotone on T, $f+h$ is lower bounded relative to S. Thus Lemma 5.1 applies and there exists an isotone submodular extension g of $f+h$ to S. Then $g-h$ is submodular on S and $g-h=f$ on T; hence $g-h$ is a submodular extension of f to S.

THEOREM 5.2. *If T is a compact sublattice of E^n and f is submodular on T and continuously differentiable on a compact convex set containing T, then there exists a submodular extension of f to E^n.*

Proof. By Taylor's theorem and the above hypotheses, there exists $M>0$ such that

$$f(y) \leqq f(x) + M \textstyle\sum_{j=1}^{n} |x_j - y_j| \tag{5}$$

for each x and y in T. Define $g(x, y) = f(x) + M \sum_{j=1}^{n} |x_j - y_j|$ for $(x, y) \in T \times E^n$. Define $h(y) = \min_{x \in T} g(x, y)$ for $y \in E^n$. Note that h exists because $g(x, y)$ is continuous in x on the compact set T for each y. Furthermore, h is submodular on E^n by Theorem 4.3 because g is submodular on $T \times E^n$. For $y \in T$, it follows from (5) that $h(y) = g(y, y) = f(y)$. Thus h is a submodular extension of f to E^n.

The extension results of Theorems 5.1 and 5.2 will have application to the isotonicity results of Section 6. The conditions of Theorem 6.1 may require that $f(x, t)$ has certain properties on a set larger than its natural domain $\bigcup_{t \in T} (S_t, t)$. However, Theorems 5.1 and 5.2 indicate that the required properties might already be implicit in the properties on $\bigcup_{t \in T} (S_t, t)$. For instance, the acyclic network example could not formally satisfy the conditions of Theorem 6.1. This is because $c(i, j)$ is only defined on $T = \{(i, j) : 1 \leqq i,\ i \leqq j-1,\ j \leqq N,\ i \text{ and } j \text{ integer}\}$. However, if $c(i, j)$ is submodular on T then it follows from Theorem 5.1 that there exists an extension of $c(i, j)$ to a submodular function on $S = \{(i, j) : 1 \leqq i \leqq N-1,\ 2 \leqq j \leqq N,\ i \text{ and } j \text{ integer}\}$; hence Theorem 6.1 is then applicable.

6. ISOTONE OPTIMAL SOLUTIONS

Consider the collection of optimization problems (1), where both the constraint set and the objective function depend upon the parameter t

for $t \in T$. Let S_t^* be the set of optimal solutions for (1) given $t \in T$, and let $T^* = \{t: S_t^* \text{ is nonempty}\}$.

We now give conditions for S_t^* to be ascending and for the selection of an element from each S_t^* so that the element is isotone in t.

LEMMA 6.1. *If S is a lattice, T is a poset, $S_t \subseteq S$ is ascending in t on T, and*

$$f(x \wedge y, t) + f(x \vee y, b) \leqq f(x, t) + f(y, b) \tag{6}$$

for all t and b in T with $t \leqq b$, $x \in S_t$, and $y \in S_b$, then S_t^* is ascending in t on T^*.

Proof. By (6) and Theorem 4.1, each S_t^* is a sublattice of S. Pick t and b in T^* with $t < b$. Then pick $x \in S_t^*$ and $y \in S_b^*$. Because $S_t \leqq^P S_t$, $x \wedge y \in S_t$ and $x \vee y \in S_b$. Then by (6) and the optimality of x and y,

$$0 \leqq f(x \vee y, b) - f(y, b) \leqq f(x, t) - f(x \wedge y, t) \leqq 0. \tag{7}$$

Thus equality holds throughout in (7) and so $x \wedge y \in S_t^*$ and $x \vee y \in S_b^*$.

Condition (6) clearly holds if $f(x, t)$ is submodular in (x, t), but then this requires the unnecessary condition that $f(x, t)$ be submodular in t for fixed x.

THEOREM 6.1. *If S is a lattice, T is a poset, $S_t \subseteq S$ is ascending in t on T, $f(x, t)$ is submodular in x on S for each $t \in T$, and $f(x, t)$ has antitone differences in (x, t) on $S \times T$, then S_t^* is ascending in t on T^*.*

Proof. By Lemma 6.1 it suffices to show that (6) holds. Pick t and b in T with $t \leqq b$, $x \in S_t$ and $y \in S_b$. Then by hypothesis $f(x \wedge y, t) - f(x, t) \leqq f(y, t) - f(x \vee y, t) \leqq f(y, b) - f(x \vee y, b)$.

Veinott (personal communication) earlier proved an isotonicity result that is a special case of Theorem 6.1. His result was for the case where each minimization problem of (1) satisfied the necessary and sufficient optimality conditions of Kuhn and Tucker [11], and thus his additional hypotheses included those of convexity, differentiability, and other regularity conditions in E^n.

The next result is an immediate consequence of Corollary 4.1, Theorem 2.3, and Theorem 6.1.

THEOREM 6.2. *If, in addition to the hypotheses of Theorem 6.1, each S_t is compact in a topology finer than the interval topology and $f(x, t)$ is lower semicontinuous in x on S_t for each $t \in T$, then each S_t^* has a least element, s_t, and a greatest element, $\bar{s}_t$, and s_t and $\bar{s}_t$ are both isotone in t on T.*

The mapping S_t^* is *strongly ascending* on T^* if t and b in T^* with $t < b$, $x \in S_t^*$ and $y \in S_b^*$ imply that $x \leqq y$. Theorem 6.3 shows that if the hypothesis of antitone differences in Theorem 6.1 is strengthened to strictly antitone differences, then S_t^* is strongly ascending.

THEOREM 6.3. *If S is a lattice, T is a poset, $S_t \subseteq S$ is ascending in t on T, $f(x, t)$ is submodular in x on S for each $t \in T$, and $f(x, t)$ has strictly antitone differences in (x, t) on $S \times T$, then S_t^* is strongly ascending in t on T^*.*

Proof. Pick t and b in T with $t<b$, $x \in S_t^*$, and $y \in S_b^*$. Suppose it is not true that $x \leqq y$. Then $y < x \vee y$ and so by hypothesis $0 \leqq f(x \wedge y, t) - f(x, t) \leqq f(y, t) - f(x \vee y, t) < f(y, b) - f(x \vee y, b) \leqq 0$, which is a contradiction.

After Theorem 6.3 appeared in an earlier version of this paper [22], the author heard from Veinott (personal communication) that he had developed a version of this result.

The isotonicity results of this section make sense intuitively when viewed in the previously discussed context of a system of complementary products. Then in such a system the optimal levels of any subset of the products are isotone as a function of the other products. While a strictly isotone transformation of $f(x, t)$ may not preserve properties like submodularity or antitone differences, such a transformation certainly preserves the isotonicity of the optimal solutions.

We will now give several brief applications of Theorems 6.1 and 6.2.

Consider a system with n products whose levels are denoted $x=(x_1, \cdots, x_n)$. The product levels x must be contained in a set S. The utility of x is $U(x)$. To acquire (or produce) each product j there is a unit price p_j, and the price vector is $p=(p_1, \cdots, p_n)$. The problem is to maximize the net value $U(x) - p \cdot x$, the utility minus the acquisition cost, over x in S. Suppose that S is a compact sublattice of E^n and $-U(x)$ is lower semicontinuous and submodular on S. The products thus are complementary, because by Theorem 3.1 the marginal utility from additional units of any given product is a nondecreasing function of the levels of each of the other products. Define $t=-p$, so the decision problem is to minimize $-U(x) - t \cdot x$ subject to x in S. Theorem 6.2 applies directly and shows that there exist optimal solutions which are an isotone function of t and hence an antitone function of p. Thus increasing the price of one product will not lead to an increase in the optimal levels of any product, and this is another reasonable definition of complementary products. Let $f(p)$ denote the optimal value of the problem of maximizing $U(x) - p \cdot x$ over x in S. This function is convex because it is the maximum of a collection of affine functions. Also, $-f(p)$ is submodular by Theorem 4.3 and the substitution $t=-p$. Thus a marginal increase in any p_j is relatively less undesirable at higher initial levels of p.

Now consider again the problem of finding minimum cost paths from one node to all other nodes in an acyclic network. Let $i(j)$ be the largest (or smallest) i optimal for j in (2). If $c(i, j)$ is submodular in (i, j) on the sublattice $\{(i, j): 1 \leq i,\ i \leq j-1,\ j \leq N,\ i \text{ and } j \text{ integer}\}$, then $i(j)$ is isotone in j. Then the recursion

$$f(j) = \min_{i(j-1) \leq i \leq j-1} (f(i) + c(i, j)), \qquad f(1) = 0$$

can replace (2). It also follows from this that if $i(j) = j-1$ for some j, then for each $k \geq j$ a minimum cost path from node 1 to node k goes through node $j-1$. Wagner and Whitin [26] established a special case in which the particular structure of their $c(i, j)$, arising from an application to inventory theory, implied that it was submodular. Eppen, Gould, and Pashigian [6] established another special case (when viewed in the *a priori* form where the conditions do not depend on $i(j)$) by finding more general conditions than those in [26] under which the inventory problem structure implied that $c(i, j)$ is submodular. Zangwill [27] also established closely related inventory applications.

Erlenkotter [7] has shown that there exist optimal solutions to the dual of the transportation problem that are isotone in the demands and antitone in the supplies, and this result also follows from Theorem 6.2.

It is possible to develop from the results of this paper Karlin's result [10] that optimal inventory policies increase as the demands increase stochastically.

In [20] the author used these isotonicity results together with Tarski's fixed-point theorem [18] to establish conditions for the existence of an equilibrium point in an n-person, nonzero sum game. That paper also gives two algorithms that, based on these results, generate a sequence of joint strategies converging monotonically to an equilibrium point.

7. SUBMODULARITY AND CONVEXITY

There are some qualitative analogies between submodular functions on lattices and convex functions on convex sets. Both submodularity and convexity are second-order properties in the sense that for twice-differentiable functions on open sets in E^n each property is equivalent to certain conditions on the matrix of second partial derivatives. The exact statements of Theorems 4.1, 4.3, and 5.2 hold for convex functions on convex sets, and there are results corresponding to Theorem 3.3, Table I, and Theorem 4.2 [5, 13]. There are also some obvious minor correspondences. This analogy makes some properties of submodular functions seem a bit more familiar and it was useful in finding certain properties of submodular functions and in developing some of the proofs. Of course, this can only be carried so far. Since every subset of E^1 is a sublattice and every real-valued function on E^1 is submodular, sublattices do not share the topological properties of convex sets and submodular functions do not have the continuity and differentiability properties of convex functions.

ACKNOWLEDGMENTS

The content of this paper is largely that of Section 1 of Chapter III of my doctoral dissertation [19]. I acknowledge the criticisms and advice of my dissertation advisor, Professor Arthur F. Veinott, Jr. This research was supported in part by the National Science Foundation and the Office of Naval Research.

REFERENCES

1. K. Baker and A. Pixley, "Polynomial Interpolation and the Chinese Remainder Theorem," *Mathematishe Z.* **143**, 165–174, (1975).
2. G. Bergman, "On the Existence of Subalgebras of Direct Products with Prescribed d-Fold Projections," *Algebra Universalis* **7**, 341–356 (1977).
3. G. Birkhoff, "A New Interval Topology for Dually Directed Sets," *Rev. Univ. Nac. Tucuman, Ser. A* **14**, 325–331 (1962).
4. G. Birkhoff, *Lattice Theory*, 3rd ed., American Mathematical Society Colloquium Publications, **25**, Providence, Rhode Island, 1967.
5. G. Dantzig, "Linear Programming under Uncertainty," *Management Sci.* **1**, 197–206 (1955).
6. G. Eppen, F. Gould, and B. Pashigian, "Extensions of the Planning Horizon Theorem in the Dynamic Lot Size Model," *Management Sci.* **15**, 268–277 (1969).
7. D. Erlenkotter, "Preinvestment Planning for Capacity Expansion: A Multi-Location Dynamic Model," doctoral dissertation, Stanford University, 1969.
8. L. Ford and D. Fulkerson, *Flows in Networks*, Princeton University Press, Princeton, 1962.
9. O. Frink, "Topology in Lattices," *Transactions of the American Mathematical Society* **51**, 569–582 (1942).
10. S. Karlin, "Dynamic Inventory Policy with Varying Stochastic Demands," *Management Sci.*, **6**, 231–258 (1960).
11. H. Kuhn and A. Tucker, "Nonlinear Programming," in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pp. 481–492, J. Neyman (Ed.), University of California Press, Berkeley and Los Angeles, 1951.
12. O. Ore, *Theory of Graphs*, American Mathematical Society Colloquium Publications, Vol. **38**, Providence, Rhode Island, 1962.
13. R. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, 1970.
14. H. Royden, *Real Analysis*, MacMillan Company, New York, 1963.
15. P. Samuelson, *Foundations of Economic Analysis*, Atheneum, New York, 1947.
16. L. Shapley, "On Network Flow Functions," *Naval Res. Logist. Quart.* **8**, 154–158 (1961).
17. L. Shapley, "Complements and Substitutes in the Optimal Assignment Problem," *Naval Res. Logist. Quart.* **9**, 45–48 (1962).
18. A. Tarski, "A Lattice-Theoretical Fixpoint Theorem and its Applications," *Pacific J. Math.* **5**, 285–309 (1955).

19. D. Topkis, "Ordered Optimal Solutions," doctoral dissertation, Stanford University, 1968.
20. D. Topkis, "Equilibrium Points in Nonzero-Sum n-Person Subadditive Games," technical report ORC 70-38, Operations Research Center, University of California, Berkeley, 1970.
21. D. Topkis, "Monotone Minimum Node-Cuts in Capacitated Networks," technical report ORC 70-39, Operations Research Center, University of California, Berkeley, 1970.
22. D. Topkis, "Minimizing a Subadditive Function on a Lattice," technical report, Hebrew University, Jerusalem, 1975.
23. D. Topkis, "Applications of Minimizing a Subadditive Function on a Lattice," unpublished paper, 1976.
24. D. Topkis, "The Structure of Sublattices of the Product of n Lattices," *Pacific J. Math.* **65,** 525–532 (1976).
25. D. Topkis, "Topology and Subcomplete Sublattices," unpublished paper, 1977.
26. H. Wagner and T. Whitin, "Dynamic Versions of the Economic Lot Size Model," *Management Sci.* **5,** 89–96 (1958).
27. W. Zangwill, "A Backlogging Model and a Multi-Echelon Model of a Dynamic Economic Lot Size Production System," *Management Sci.* **15,** 506–527 (1969).