Mathematical Programming. The State of the Art, Springer 1983, 235–257.

(235)

QT2:8227X3XX46.0S

Submodular functions and convexity

L. Lovász
Eötvös Loránd University, Department of Analysis I, Múzeum krt. 6-8, H-1088
Budapest, Hungary

0. Introduction

In "continuous" optimization convex functions play a central role. Besides elementary tools like differentiation, various methods for finding the minimum of a convex function constitute the main body of nonlinear optimization. But even linear programming may be viewed as the optimization of very special (linear) objective functions over very special convex domains (polyhedra). There are several reasons for this popularity of convex functions:

- Convex functions occur in many mathematical models in economy, engineering, and other sciencies. Convexity is a very natural property of various functions and domains occuring in such models; quite often the only non-trivial property which can be stated in general.

- Convexity is preserved under many natural operations and transformations, and thereby the effective range of results can be extended, elegant proof techniques can be developed as well as unforeseen applications of certain results can be given.

- Convex functions and domains exhibit sufficient structure so that a mathematically beautiful and practically useful theory can be developed.

- There are theoretically and practically (reasonably) efficient methods to find the minimum of a convex function.

It is less apperant, but we claim and hope to prove to a certain extent, that a similar role is played in discrete optimization by *submodular set-functions*. These functions do not enjoy the nice geometric image of convex functions, and accordingly their significance has been discovered only gradually. The firs class of submodular functions which was studied thoroughly was the class of matroid rank functions. Generalizing certain basic properties of matroid polyhedra, Edmonds (1970) began the systematical study of submodularity. Let us remark, however, that approaching from quite a different angle, Choquet (1955) also introduced these set-functions. He proved that the newtonian "capacity" of a subset of $\mathbb{R}^3$ defines a submodular set-function. Quite a few proof techniques in graph theory, but also in probability, geometry and lattice theory have made implicit use of submodularity of certain set-functions, thus forecasting the importance of this property.

In this paper we survey some of the most important aspects of submodularity. In particular, we shall see that submodularity shares the above-listed four

24. MRZ. 1983

important properties of convexity. But besides this formal analogy, we shall develop a more fundamental connection between these two concepts. This connection will enable us to apply some basic facts concerning convex functions to obtain similarly basic results for submodular functions. In particular, a polynomial-time algorithm to find the minimum of a submodular set-function and a "sandwich theorem" for submodular functions can be obtained this way. (A refined integral version of this sandwich theorem, due to A. Frank (1980), needs a direct proof; but this powerful result is also motivated clearly by the convex-submodular analogy.)

Somewhat surprisingly, submodular functions are in some respects also similar to concave functions. This suggests that the maximization problem for submodular functions may also lead to interesting results. We shall see that this problem is substantially more difficult than the minimization problem, and in general no solution exists; but solutions for special cases, as well as reasonable heuristics, can be obtained.

Submodularity gives rise to polyhedra with very nice properties. Following Edmonds (1970), we treat various linear programming problems associated with submodular functions. For a single submodular function, these linear programming problems can be solved by appropriate versions of the *greedy algorithm*. For two submodular functions similar polyhedral considerations lead to very deep min-max results which can be examplified by the Matroid Intersection Theorem.

Several recent combinatorial studies involving submodularity fit into the following pattern. Take a classical graph-theoretical result (e.g. the Marriage Theorem, the Max-flow-min-cut Theorem etc.), and replace certain linear functions occuring in the problem (either in the objective function or in the constraints) by submodular functions. Often the generalizations of the original theorems obtained this way remain valid; sometimes even the proofs carry over. What is important here to realize is that these generalizations are by no means l'art pour l'art. In fact, the range of applicability of certain methods can be extended enormously by this trick. Choosing the submodular functions from among the many submodular functions arising from graphs and other combinatorial structures, various and often surprising results can be obtained this way. As an example, one can obtain, as a "submodular" generalization of the Marriage Theorem, a version of the famous Matroid Intersection Theorem, which in turn is known to imply Menger's Theorem, the Disjoint Spanning Tree Theorem and many other graph-theoretical results.

These are the main aspects of submodularity this paper will survey. We shall concentrate on some of the fundamental ideas and constructions; only a few proof will be described in detail, in cases when no appropriate reference can be given. This paper cannot undertake the task of a comprehensive survey of all aspects of submodularity, in particular if we consider matroid theory as a special case; for this we refer the interested reader to Welsh (1976), and shall assume that the reader is at least in part familiar with it. A forthcoming book of A. Schrijver is also strongly recommended for further reading on related subjects.

24. MRZ. 1983

1. Definitions and examples

Let S be a finite set and $f: 2^S \to \mathbb{R}$ a real valued function defined on the subsets of S. The set-function f is called *submodular* if the following inequality holds for any two subsets X and Y of S:

$$f(X \cap Y) + f(X \cup Y) \leqslant f(X) + f(Y). \tag{1}$$

The setfunction f is called *supermodular* if the reversed inequality holds true for every pair of subsets. Finally, f is called *modular* if it is both submodular and supermodular, i.e. the following equality holds identically for any two subsets X and Y of S:

$$f(X \cap Y) + f(X \cup Y) = f(X) + f(Y). \tag{2}$$

Let us remark immediately that modular functions are very simple. In fact, (2) implies easily by induction that

$$f(X) = f(\emptyset) + \sum_{x \in X} f(\{x\}). \tag{3}$$

Conversely, if $f(\emptyset)$ and $f(\{x\})$ $(x \in S)$ are prescribed arbitrarily, then the set-function defined by (3) is modular. If we identify every subset of S with its indidence vector, then modular functions will correspond to linear functions (restricted to 01-vectors).

Submodular (and supermodular) set-functions are much more interesting then modular ones, and there are many highly non-trivial examples of such set-functions. To verify submodularity of a set-function, the following characterization of submodularity often helps.

1.1. Proposition. Let f be a set-function defined on all subsets of S. Then f is submodular if and only if the derived set-functions

$$f_a(X) = f(X \cup \{a\}) - f(X) \qquad (X \subseteq S - \{a\})$$

are monotone decreasing for all $a \in S$.

Obviously, an analogous characterization of supermodular set-functions can be formulated, by replacing "decreasing" by "increasing".

We continue with a number of examples of submodular and supermodular set-function. As mentioned in the introduction, the main aim of these examples is to support the thesis that submodularity shares the first important property of convexity, namely that submodular functions occur in many mathematical models in a very natural way.

1.2. Example. Let S be the set of columns of a matrix A and let, for each $X \subseteq S$, $r(X)$ be the rank of the matrix formed by the columns in X. Then r is submodular. This fact follows by elementary linear algebra. For the reader familiar with matroid theory it is clear at this point that this example is just a special case of matroid rank functions, and so many more examples could be manufactured

24. MRZ. 1983

by taking other classes of matroids (graphic, transversal etc.). Some of these will occur below in different contexts.

1.3. Example. Let G be a (directed or undirected) graph and $X \subseteq V(G)$. Let $\delta(X)$ denote the number of lines of G connecting X to $V(G)-X$. Then a simple counting argument shows that $\delta(X)$ is a submodular setfunction on the subsets of $S=V(G)$. This fact has been the cornerstone of many studies in graph theory concerning connectivity, flows, and other problems. We can generalize this construction as follows. Let $w: E(G) \to \mathbb{R}_+$ be any weighting of the lines of G with non-negative real numbers, and define $\delta_w(X)$ as the sum of weights of the lines of G connecting X to $V(G)-X$. Instead of graphs, we could also consider hypergraphs.

1.4. Example. Let G be a bipartite graph, with bipartition $\{A, B\}$. For each $X \subseteq A$, let $\Gamma(X)$ denote the set of neighbors of X. Then $|\Gamma(X)|$ is a submodular set-function on the subsets of A. The submodularity of this set-function plays role in some treatments of the bipartite matching problem, see e.g. Ore (1955). From $|\Gamma(X)|$ we can construct another submodular set-function on the subsets of A by the formula

$$\nu(X)=|X|+\min\{|(Y)|-|Y|: Y \subseteq X\}.$$

From the König-Hall Theorem we cann see that $\nu(X)$ gives the maximum number of nodes in X which can be matched with nodes in B by disjoint lines of G. Thus ν is in fact the rank function of the transversal matroid on A defined by the bipartite graph G.

1.5. Example. Let G be a graph. For each $X \subseteq E(G)$, let $c(X)$ denote the number of connected components of the subgraph $(V(G), X)$. Then $c(X)$ is a supermodular set-function, which can be easily verified e.g. by using Proposition 1.1. If one wishes to see a submodular function, one may consider $-c(X)$ or (if one wants it non-negative) one may consider $|V(G)|-c(X)$. This function is then the rank function of the circuit matroid of the graph G.

1.6. Example. We may modify the previous example as follows. For each $X \subseteq V(G)$, let $d(X)$ denote the number of connected components of the graph $G-X$. Then $d(X)$ is in general neither submodular nor supermodular. But if we restrict its domain to the subsets X of an independent set A of points of G, then $d(X)$ will be supermodular.

1.7. Example. Let B be a bar-and-joint structure, i.e. a graph whose nodes are points of the euclidean 3-space, and whose lines are considered rigid bars which are attached to the nodes by flexible joints. Let, for each $X \subseteq V(G)$, $f(X)$ denote the degree of freedom of the subset X, i.e. the dimension of the infinitesimal motions of the nodes in X which extend to an infinitesimal motion of all nodes compatible with the given bars. Thus, $f(\emptyset)=0$, $f(\{x\})=3$ for every $x \in V(G)$, and $f(\{x, y\})=5$ or 6 depending on whether or not the whole structure forces x and y to stay at the same distance, etc. It follows from the ele-

24. MRZ. 1983

ments of the theory of rigid bar-and-joint structures that f is a submodular set-function of the subsets of $V(G)$. See Crapo (1979).

1.8. Example. Let $A_1, \dots, A_n$ be random events, $S = \{A_1, \dots, A_n\}$, and let, for $X = \{A_{i_1}, \dots, A_{i_k}\} \subseteq S$,

$$f(X) = \mathrm{Prob}(A_{i_1} \dots A_{i_k})$$

denote the probability that all events in X occur. Then f is a supermodular set-function on the subsets of S. This fact is used implicity in some sieving techniques.

1.9. Example. Let G and H be two graphs, $X \subseteq (G)$ and $Y \subseteq E(H)$. Let $h(X)$ denote the number of homomorphisms of the graph $(V(G), X)$ into H, and let $g(Y)$ denote the number of homomorphisms of the graph G into the graph $(V(H), Y)$. Then g and h are supermodular set-functions. This fact (which can be verified easily using Proposition 1.1) can be generalized in many ways to categories etc., becoming more and more trivial. (This last remark shows that submodularity is not a deep property, but it is often difficult to recognize the circumstances under which it occurs.)

1.10. Example. This last example comes from geometry. Let $a_1, \dots, a_n$ be linearly independent vectors in $\mathbb{R}^n$, and let, for $X = \{a_{i_1}, \dots, a_{i_k}\}$,

$$f(X) = \log \mathrm{vol}_k (a_{i_1}, \dots, a_{i_k}),$$

where vol_k denotes the k-dimensional volume of the parallelepiped spanned by $a_{i_1}, \dots, a_{i_k}$. Then

$$\begin{aligned} f_a(X) &= \log \mathrm{vol}_{k+1}(a_{i_1}, \dots, a_{i_k}, a) - \log \mathrm{vol}_k (a_{i_1}, \dots, a_{i_k}) \\ &= \log \frac{\mathrm{vol}_{k+1}(a_{i_1}, \dots, a_{i_k}, a)}{\mathrm{vol}_k (a_{i_1}, \dots, a_{i_k})} \end{aligned}$$

is the logarithm of the "height" of a above the plane spanned by $a_{i_1}, \dots, a_{i_k}$. Hence Proposition 1.1 easily implies that f is submodular.

The first five examples above are well-understood and well-studied submodular functions. The last four examples show that submodularity arises in geometry, probability and other branches of mathematics and one could also mention sporadic use of these examples. Nevertheless, these examples and in particular the significance of submodularity are much less understood. One may hope that an in-depth study of submodularity will have some contribution to "classical" mathematical problems as well.

Often it is useful to extend the definition of submodularity and supermodularity to partial set-functions, i.e. to functions which are only defined on a family $\mathscr{H}$ of subsets of S. Then of course (1) is required to hold only if all four of $X, Y, X \cap Y$, and $X \cup Y$ belong to $\mathscr{H}$.

If we study submodularity of a set-functioned defined only on a family $\mathscr{H} \subseteq 2^S$, then it is natural to assume that there are "fairly many" pairs X, Y for which (1) is defined (and therefore is assumed to hold). There are three types of

24. MRZ. 1983
1. Hauptkontakt

set-systems $\mathscr{H}$ defined by this kind of condition which have been studied extensively. The family $\mathscr{H} \subseteq 2^S$ is called a *lattice family* if X, Y implies that $X \cap Y \in \mathscr{H}$ and $X \cup Y \in \mathscr{H}$. $\mathscr{H}$ is called an *intersecting family* if $X, Y \in \mathscr{H}$, $X \cap Y \neq \emptyset$ implies that $X \cap Y \in \mathscr{H}$ and $X \cup Y \in \mathscr{H}$. Finally, $\mathscr{H}$ is called a *crossing family* if $X, Y \in \mathscr{H}$, $X \cap Y \neq \emptyset$, $X \cup Y \neq S$ implies that $X \cap Y \in \mathscr{H}$ and $X \cup Y \in \mathscr{H}$. For an intersecting family we shall assume that $\emptyset \notin \mathscr{H}$, and for a crossing family we shall assume that $\emptyset \notin \mathscr{H}$ and $S \notin \mathscr{H}$. Many results on submodular set-functions defined on all subsets of a set extend without any change to submodular set-functions defined on lattice families, and with natural modifications to submodular set-functions defined on intersecting or crossing families.

Let us mention two examples of submodular functions which are defined on subfamilies of 2^S only.

DT2:8227X3XX47.0S

1.11. Example. Let G be a directed graph, $S = V(G)$, and let $\mathscr{H}$ be the family of those subsets of S which determine a directed cut, i.e. those subsets X for which no line of G has head in X and tail in $S - X$. Then the function $\delta(X)$ introduced in Example 1.3, when restricted to $\mathscr{H}$, will not only be submodular but even modular. Further, $\mathscr{H}$ is a lattice family.

1.12. Example. The most common intersecting and crossing families are the families of all non-empty subsets and of all non-empty proper subsets of S, respectively. It is easy to see that every submodular set-function defined on 2^S, when restricted to the non-empty subsets of S, gives a submodular set-function on this latter family and conversely, every submodular set-function defined on the non-empty subsets of S can be extended to a submodular set-function defined on all subsets (by letting $f(\emptyset)$ be a very small number). However, if we restrict our attention to, say, non-negative submodular functions then we obtain essentially more general problems if we consider set-functions defined on the non-empty subsets of S only. Similar remarks apply to the family of non-empty proper subsets.

We close this section with a few more definitions. A submodular function which is integral valued, monotone increasing, and has value 0 on $\emptyset$, is called a *polymatroid function*. The submodular function arising from a bar-and-joint structure (Example 1.7) is a polymatroid function. Further, every matroid rank function is a polymatroid function. Let us see one more example.

1.8. Example. Let S be a set of subspaces of a linear space and let, for each $X \subseteq S, f(X)$ denote the dimension of the linear subspace generated by the subspaces in X. Then f is a polymatroid function. More generally, let (E, r) be a matroid (r being its rank function), and let $S \subseteq 2^E$. For each $X \subseteq S$, let

$$f(X) = r(\cup X).$$

Then f is a polymatroid function. It will follow from the discussions in the next section that every polymatroid function, in fact, arises by this construction from some matroid (Helgason 1974, McDiarmid 1975).

24 MRZ. 1983
1. Hauptschalter

A set-function f defined on the subsets of S will be called *t-smooth*, if

$$|f(X\cup\{a\})-f(X)|\leq t$$

for every $a\in S$ and $X\subseteq S-\{a\}$. This is equivalent to saying that

$$|f(X)-f(Y)|\leq t|X\,\Delta\, Y|$$

holds for all X, $Y\subseteq S$. Clearly, 1-smooth polymatroid functions are just matroid rank functions.

2. Operations on submodular functions

Following our line of establishing submodular set-functions as discrete analogs of convex functions, we turn to the second property of convexity stated in the introduction, namely that there are many natural operations which manufacture new submodular functions from old ones. Such operations are useful to extend the effective range of certain theorems (by applying them to modified functions), to prove submodularity (by recognizing that the function arises by a known construction from known submodular functions), to construct examples etc. But as it turns out, the study of various properties of some of these constructions yields sufficient insight into submodularity so that the study of these constructions may also be considered as the structure theory of submodular functions. So the results to be discussed in this section should also prove that submodular set-functions have sufficient structure so that a mathematically beautiful and practically useful theory can be developed.

Let us start with the trivialities. If f is submodular then the set-function $f(S-X)$ is also submodular, while $-f(X)$ is supermodular. If f and g are submodular then $f+g$ is also submodular.

Let f_i be a submodular set-function defined on the subsets of S_i $(i=1,\dots,k)$, where the S_i' are mutually disjoint. Then the *direct sum* of $f_1,\dots,f_k$ is the set-function f defined on $S=S_1\cup\dots\cup S_k$ by the formula

$$f(X)=f_1(S_1\cap X)+\dots+f_k(S_k\cap X).$$

The following construction makes a monotone function from a non-monotone; cf. example 1.4. If f is any set-function defined on the subsets of S, we set

$$f_{\mathrm{mon}}(X)=\min\{f(Y)\colon Y\subseteq X\}.$$

Clearly, f_{mon} is monotone decreasing. The following is slightly less obvious:

2.1. Proposition. If f is a submodular set-function then f_{mon} is also submodular.

Neither the minimum nor the maximum of two submodular set-functions is in general submodular. But the following fact is useful in many constructions:

24. MRZ. 1983
1. Kauckenpaktor

2.2. Proposition. Let f and g be submodular set-functions such that $f-g$ is either monotone increasing or monotone decreasing. Then $\min(f,g)$ is also submodular.

We now come to a very important and non-trivial construction. Let f be any set-function defined on the subsets of X. Define the *Dilworth truncation of f* as the set-function

$$f_*(\emptyset) = 0,$$
$$f_*(X) = \min\{f(X_1)+\ldots+f(X_k)\colon X_1\cup\ldots\cup X_k = X,\ X_i\neq\emptyset,\ X_i\cap X_j=\emptyset\}$$

This construction generalizes the notion of Dilworth truncation of matroids; hence the name (Dunstan 1976). For properties and geometric background, see also Mason (1977, 1981) and Lovász (1977).

2.3. Proposition. If f is a submodular set-function then its Dilworth truncation f_* is also submodular.

Note that if $f(\emptyset)\geqq 0$ then $f_*=f$ on the non-empty subsets of S. In general, the relationship between f_* and f is somewhat more complicated:

2.4. Proposition. Let f be a submodular set-function defined on the subsets of S. Then f_* is the unique maximal set-function among all set-functions g with the following properties:

a) g is submodular,
b) $g(\emptyset)=0$,
c) $g(X)\leqq f(X)$ for all non-empty $X\subseteq S$.

We can define the "upper Dilworth truncation" f^* for any set-function by replacing "min" by "max" in the definition of Dilworth truncation. Then we find that upper Dilworth truncation preserves supermodularity.

Let f and g be set-functions defined on the subsets of S. We defined their *convolution* $h=f*g$ by

$$h(X)=\min\{f(Y)+g(X-Y)\colon Y\subseteq X\}.$$

The reader familiar with the Matroid Intersection Theorem will notice that if f and g are two matroid rank functions then h is the rank function of their intersection. Since the intersection of two matroids is not a matroid in general, it follows that convolution does not always preserve submodularity. However, the following is true.

2.5. Theorem. The convolution of a submodular and a modular set-function is submodular.

Proof. Let f be a submodular, g a modular set-function, both defined on the subsets of S, and let $h=f*g$ be their convolution. Let $X,Y\subseteq S$, we claim that

$$h(X)+h(Y)\geqq h(X\cap Y)+h(X\cup Y). \tag{4}$$

By definition, there exist subsets $U\subseteq X$ and $V\subseteq Y$ such that

$$h(X)=f(U)+g(X-U),\quad h(Y)=f(V)+g(Y-V).$$

24. MRZ. 1983

Thus

$$\begin{aligned} h(X)+h(Y) &= f(U)+g(X-U)+f(V)+g(Y-V) \\ &\geq f(U\cap V)+f(U\cup V)+g((X-U)\cap(Y-V)) \\ &\quad +g((X-U)\cup(Y-V)). \end{aligned}$$

However, $(X-U)\cap(Y-V)=((X\cap Y)-(U\cap V))\cap((X\cup Y)-(U\cup V))$ and $(X-Y)\cup(Y-V)=((X\cap Y)-(U\cap V))\cup((X\cup Y)-(U\cup V))$, and so using the modularity of g,

$$\begin{aligned} h(X)+h(Y) &\geq f(U\cap V)+f(U\cup V)+g((X\cap Y)-(U\cap V)) \\ &\quad +f((X\cup Y)-(U\cup V)). \end{aligned}$$

But here we have, by the definition of h,

$$f(U\cap V)+g((X\cap Y)-(U\cap V))\geq h(X\cap Y)$$

and

$$f(U\cup V)+g((X\cup Y)-(U\cup V))\geq h(X\cup Y),$$

and so (4) follows.

Note that $f_{\mathrm{mon}}=f*0$, and so proposition 2.1 is a special case of Theorem 2.5. We shall see more significant applications below.

Theorem 2.5 implies the following description of the convolution of a submodular and a modular function:

2.6. Corollary. Let f be a submodular set-function and g a modular set-function defined on the subsets of S. Assume that $f(\emptyset)=g(\emptyset)=0$. Then $f*g$ is the unique largest submodular set-function h satisfying $h(\emptyset)=0$, $h\leq f$ and $h\leq g$.

A special case of the convolution construction which we shall use is the following. Let the modular set-function g be defined by $g(\emptyset)=0$, $g(a)=\alpha$, and $g(x)=+\infty$ if $x\in S-a$ (where $a\in S$ is a fixed element and α any real number). Then it follows easily that

$$(f*g)(X)=\min\{f(X), f(X-a)+\alpha\},$$

in particular $(f/g)(X)=f(X)$ if $a\in X$. The set-function $f*g$ will be called the *local truncation of f at the point a with value α*. If $f(\emptyset)=0$, then this local truncation is the unique largest submodular set-function h such that $h(\emptyset)=0$, $h\leq f$ and $h(a)\leq\alpha$. In particular, $h=f$ if $f(a)\leq\alpha$.

If $(S, r_1), \ldots, (S, r_k)$ are matroids on the same set S, then the rank function of their sum can be written, by the results of Edmonds-Fulkerson (1965), as

$$r=(r_1+\ldots+r_k)*\mathrm{card},$$

where $\mathrm{card}(X)=|X|$ is the cardinality function. In fact it follows from the above results that r is the rank function of a matroid, and that this is the unique free-est matroid satisfying

$$r\leq r_1+\ldots+r_k.$$

We come to the problem of extending submodular set-functions - in two different senses. First we investigate how a submodular function defined on a

24. MRZ. 1983
1.

lattice (intersecting, crossing) family of sets can be extended to all subsets. The case of a lattice family is easy. Let $\mathscr{H}$ be a lattice family of subsets of S and f a submodular function on $\mathscr{H}$. $\mathscr{H}$ contains a unique largest subset S'. For every $X \subseteq S'$, $\mathscr{H}$ contains a unique smallest subset including X; let this subset be denoted by $\bar{X}$. Then the following formula defines an extension of f to all subsets:

$$\bar{f}(X) = f(\overline{X \cap s'}).$$

The extension procedure is similar for crossing families. Let $S_1, \dots, S_k$ be the maximal members of the crossing family $\mathscr{H}$. Clearly $S_1, \dots, S_k$ are disjoint. For every $X \subseteq S_i$, $X \neq \emptyset$, there is a unique smallest set in $\mathscr{H}$ including X; let this be denoted by $\bar{X}$. For $1 \leq i \leq k$ and $X \subseteq S$, let

$$f_i(X) = \begin{cases} f(\overline{X \cap S_i}) & \text{if } X \cap S_i \neq \emptyset, \\ -c & \text{if } X \cap S_i = \emptyset. \end{cases}$$

(where c is a parameter to be chosen appropriately). Then

$$\bar{f}(X) = \sum_{i=1}^{k} f_i(X) + (k-1)c$$

is an extension of f and is submodular provided c is large enough.

In the case of a crossing family, we first extend the definition of f to those subsets $X \subseteq S$, $X \neq S$ which can be written as the union of sets in H, by the formula

$$f'(X) = \min\left\{ \sum_{i=1}^{k} f(X_i) + (k-1)c : \{X_1, \dots, X_k\} \text{ is a partition of } X \right\}$$

(where again c is a parameter). In particular,

$$f'(\emptyset) = -c.$$

The complements of these subsets form an intersecting family, and so f' can be extended to a submodular set-function defined on all subsets of S by a procedure which is "dual" to the one used in the case of intersecting families.

The main properties of these extensions are given in the following proposition:

DT2:8227X3XX48.0S

2.7. Proposition. A submodular function defined on a crossing family of subsets can be extended to a submodular set-function defined on all subsets. If the original set-function is integral valued, so is its extension. If the original set-function is non-negative, then its extension is also non-negative provided the family is lattice, and is non-negative on the non-empty subsets if the family is intersecting.

Let us consider now extension in a different sense, namely extending a submodular set-function to a larger underlying set. This is in general a very difficult problem and even in the case of matroids it cannot be considered well-solved. Crapo (1965) gave a description of single-element extensions of matroids, but his conditions are often difficult to apply. Some extension proce-

24. MRZ. 1983

dures, like principal extensions, are however well-understood and very useful tools in matroid theory.

We shall not attempt to develop a general theory of extensions of submodular functions, but shall confine ourselves with offering a few constructions.

A set-function f' defined on the subsets of S' is called an *extension* of the set-function f defined on the subsets of the set S if $S' \subseteq S$ and $f(X) = f'(X)$ for every $X \subseteq S$.

We have already defined the direct sum of submodular functions. This may also be viewed as a very simple extension procedure.

A very simple but important extension procedure in matroid theory is *parallel extension*. For submodular set-functions, this procedure can be carried out more generally – with some restrictions. Let f be a set-function defined on the subsets of S. Let $T \subseteq S$, and define a new set-function as follows. Take a new element $a_T \in S$, set $S' = S \cup \{a_T\}$, and define the *extension of f by an element a_T parallel to the subset T by*

$$f'(X) = f(X),$$
$$f'(X \cup \{a_T\}) = f(X \cup T) \quad \text{for } X \subseteq S.$$

Unfortunately, the parallel extension of a submodular set-function is not always submodular. Let an element $a \in S$ be called *increasing* with respect to the set-function f if $f(X \cup \{a\}) \geq f(X)$ for all $X \subseteq S$. If f is a monotone increasing set-function then of course every element of S is increasing with respect to f. Using this notion, the following result characterizing submodularity of parallel extensions can be proved easily.

2.7. Proposition. Let f be a set-function defined on the subsets of S, let $T \subseteq S$ and f' an extension of f by an element parallel to T. Then f' is submodular if and only if f is submodular and every element of T is increasing with respect to f.

An extension parallel to the subset T followed by a local truncation at the new element with value α is called a *principal extension on the subset T with value α*. It follows that if T consists of increasing elements and the original function is submodular, then so is this principal extension. This construction generalizes the principal extension of matroids.

We remark the following important property of principal extensions. Let f be a submodular set-function defined on the subsets of S, and let T_1 and T_2 be two subsets of S consisting of increasing elements. Let, further, α_1 and α_2 be real numbers. Construct the principal extension of f on T_1 with value α_1, and then construct the principal extension of the resulting set-function on T_2 with value α_2. This way we obtain a submodular set-function f'. Let us carry out the same construction but with the indices 1 and 2 interchanged, to get a submodular set-function f''. It is easy to verify that $f' = f''$. This fact can be expressed compactly as follows:

2.8. Proposition. Principal extensions commute.

A special case of the parallel extension has been used in example 1.8. In fact, given a matroid and a set of subsets, we may extend it by elements parallel

24. MRZ. 1983

to the given subsets, and then delete the original elements. This way we obtain a polymatroid function. Let us show now that conversely, every polymatroid function arises this way (Helgason 1974, McDiarmid 1975):

2.9. Thorem. Given any polymatroid function f defined on the subsets of a set S, there exists a matroid (E, r), and a mapping $\varphi: S \to 2^E$, such that

$$f(X) = r(\cup \varphi(X)) \tag{5}$$

holds for every $X \subseteq S$.

Proof (sketch). For every $a \in S$, construct $f(a)$ times the principal extension of f on the subset a with value 1. It follows from proposition 2.8 that the order in which this is done is irrelevant. Then delete the original elements. This way a matroid is obtained. Let, for each $a \in S$, $\varphi(a)$ denote the set of new elements added on a. Then it is easy to verify that (5) holds.

Principal extensions and Dilworth truncations have a very nice geometric meaning. Here we only sketch the background. Let S be a collection of subspaces of a linear space and let $f(X)$ denote the dimension of the subspace generated by the members of X (example 1.8).

Then a parallel extension of f by an element parallel to $T \subseteq S$ can be obtained by adding the linear span of T to S. A local truncation of f at $a \in S$ with value α (where $0 < \alpha \leq f(a)$) can be constructed by replacing a by an α-dimensional subspace of a "in general position". This latter phrase can be made precise by requiring that the new subspace is not contained in the span of any subset of S unless the whole of a is. It is not difficult to see that if the underlying field is large enough then such a "general" subspace always exists.

Combining these two observations we see that a principal extension of f on T with value α (where $0 < \alpha \leq f(\mathbf{I})$) can be obtained by adding to S a "general" α-dimensional subspace of the linear span of T.

The function $f-1$ is also submodular (and monotone), but its value is negative on $\emptyset$. It is easy to see that $(f-1)_*$ is a polymatroid function. A representation of this can be constructed from a representation of f by linear subspaces as follows. Take a "general" hyperplane h in the whole space, and replace each $a \in S$ by $a \cap h$. Here, however, the notion of "general" is more involved. The easiest way out is to extend the underlying field with infinitely many algebraically independent transcendentals, and then take a gebraically independent transcendentals as the coefficients of the equation of h. For details see Lovász (1977, 1982) and Mason (1977, 1981).

3. Polyhedra associated with submodular functions

Let f be a submodular set-function defined on the subsets of S and let $f(\emptyset) = 0$. Consider the following polyhedron:

$$P_f = \{x \in \mathbb{R}^S : x(T) \leq f(T) \quad \text{for all } T \subseteq S\}.$$

24. MRZ. 1983
1. ...

(Here, as usual $x(T)=\sum_{t\in T} x(t)$.) We shall be concerned with finding the maximum of a linear objective function $c\cdot x$ over $x\in P_f$. To exclude trivial cases, assume that $c\geqq 0$.

Note that this problem is a linear programming problem and that it has a finite optimum solution. Therefore we may consider its dual and this will have the same optimum value. We shall have a variable y_T for every $T\subseteq S$, and the constraints

$$y_T\geqq 0,$$

$$\sum_{T\ni t} y_T=c(t) \quad \text{for all } t\in S.$$

The dual objective function is

$$\sum_T f(T)y_T$$

f.

The fact that this is a linear program, however, is of little help here since the number of constraints in the primal program (and hence the number of variables in the dual) is exponentially large. Both problems can be solved however by a very simple procedure called the *greedy algorithm*.

To formulate this, let us first label the elements of $S=\{a_1,\dots,a_n\}$ so that $c(a_1)\geqq c(a_2)\geqq \dots \geqq c(a_n)$. Then an optimal primal solution is given by

$$x(a_k)=f(\{a_1,\dots,a_k\})-f(\{a_1,\dots,a_{k-1}\}) \quad (1\leqq k\leqq n). \tag{6}$$

Furthermore, an optimum dual solution is given by

$$y_T=\begin{cases} c(a_k)-c(a_{k+1}), & \text{if } T=\{a_1,\dots,a_k\} \text{ for some } k,\ 1\leqq k\leqq n, \\ 0, & \text{otherwise}. \end{cases} \tag{7}$$

c f

Why are these solutions called greedy? An explanation is given by the following interpretations. [redacted] can arrive at the primal solution described above as follows. First, choose the value of the variable $x(a_1)$ as large as possible. There is a trivial upper bound on $x(a_1)$, namely $f(\{a_1\})$, so let

(2)

$$x(a_1)=f(\{a_1\}).$$

Next, choose the value of $x(a_2)$ as large as possible. There are now two trivial upper bounds: $f(\{a_2\})$ and $f(\{a_1,a_2\})-f(\{a_1\})$. From the submodularity of f it follows that the second one of these is always the smaller, so let

$$x(a_2)=f(\{a_1,a_2\})-f(\{a_1\})$$

etc. This arg ent also shows that the vector given by (6) is indeed a feasible point of P_f. (Note that optimality is not yet proved; the greedy solution is only optimal for very special linear programs.)

Second, we show how a greedy interpretation of the dual solution (7) can be given. Consider first the dual variable y_T with the largest T, namely $T=S$, and choose its value as large as possible. Clearly this largest possible value is $c(a_n)$, so fix the value

$$y_S=c(a_n).$$

24. MRZ. 1983

Once this value is fixed, it follows immediately from the constraint corresponding to $t=a_n$ that $y_T=0$ for every other subset T containing a_n. Consider the next largest subset T of S whose dual variable is not yet fixed; this is clearly $S-a_n=\{a_1,\dots,a_{n-1}\}$; and chose y_{S-a_n} as large as possible, which clearly means

$$y_{S-a_n}=c(a_{n-1})-c(a_n)$$

etc. Again, this argument proves that the greedy dual solution (7) is indeed a feasible dual solution, but does not prove optimality. However, the optimality of both the primal and the dual greedy solution follows now easily from the fact that they give the same objective value:

$$\sum_{k=1}^{n} c(a_k)(f(\{a_1,\dots,a_k\})-f(\{a_1,\dots,a_{k-1}\}))$$
$$=\sum_{k=1}^{n} f(\{a_1,\dots,a_k\})(c(a_k)-c(a_{k+1})).$$

The fact that (6) and (7) are optimal solutions has some important consequences. First note that if f is integral valued then (6) is integral and if c is integral then (7) is integral. Hence it follows:

3.1. Theorem. If f is integral valued then P_f has integral vertices. The system defining P_f is total dual integral.

A much stronger reult is due to Edmonds and Giles (1977).

3.2. Theorem. Let f and g be integral valued submodular set-functions defined on the subsets of the same set S, and assume that $f(\emptyset)=f(\emptyset)=0$. Then $P_f\cap P_g$ has integral vertices and the system

$$x(T)\leq\min\{f(T),g(T)\}$$

(which describes $P_f\cap P_g$) is total dual integral.

Many graph-theoretical and combinatorial results follow from theorem 3.2, most notably the Matroid Intersection Theorem. We shall discuss further applications in section 6.

Let $\hat{f}(c)$ denote the maximum of $c\cdot x$ subject to $x\in P_f$. Let us consider the case when c is the incidence vector of a set $T\subseteq S$. Then it follows from either the primal or the dual greedy algorithm that $\hat{f}(c)=f(T)$. Thus $\hat{f}(c)$ may be considered as an extension of the function f, which originally is defined on 01-vectors, to all non-negative vectors.

Such an extension can be defined for an arbitrary set-function. For let f be any set-function defined on the subsets of a set S. For the incidence vector a of a set $T\subseteq S$, set

$$\hat{f}(a)=f(T).$$

So $\hat{f}$ is defined for 01-vectors. Next let c be any non-negative vector. Then c can be written uniquely in the following form:

$$c=\sum_{i=1}^{k}\lambda_i a_i,$$

ST2:8227X3XX49.0S
24. MRZ. 1983
1. Hauskorrektur

where $\lambda_i > 0$ and $a_1 \geqq a_2 \geqq \ldots \geqq a_k$ are distinct 01-vectors. Define

$$\hat{f}(C) = \sum_{i=1}^{k} \lambda_i \hat{f}(a_i).$$

It is clear that this way a well-defined function is obtained.

Using this construction, we can re-formulate Theorem 3.2 is a form closely resembling the Matroid Intersection Theorem:

3.3. Theorem. Let f and g be integral valued submodular set-functions defined on the subsets of the same set S, and assume that $f(\emptyset) = g(\emptyset) = 0$. Further let $c \geqq 0$ be an integral vector. Then

$$\max\{c \cdot x : x \text{ integral}, \ x(T) \leqq f(T) \text{ for all } T \subseteq S\}$$
$$= \min\{\hat{f}(x) + \hat{g}(c - x) : x \text{ integral}, \ 0 \leqq x \leqq c\}.$$

In the special case when $c = 1$, which is most important from the combinatorial point of view, the right hand side is even simpler:

3.4. Corollary. Let f and g be integral valued submodular set-functions defined on the subsets of the same set S, and assume that $f(\emptyset) = g(\emptyset) = 0$. Then

$$\max\{1 \cdot x : x \text{ integral}, \ x(T) \leqq f(T) \text{ for all } T \subseteq S\}$$
$$= \min\{f(X) + g(S - X) : X \subseteq S\}.$$

Note that the right hand side can be written as $(fg)(S)$. In the special case when f and g are the rank functions of two matroids, the left hand side is just the cardinality of a maximum common independent set of the two matroids, and so we obtain the Matroid Intersection Theorem.

4. Submodularity and convexity

Are submodular set-functions more like convex or like concave functions? In this section we discuss some properties of them which are analogous to properties of convex functions; in the next section we shall survey some aspects of submodularity which relate it to concavity. The reader may then decide how he or she would answer the question above.

A very important connection between convexity and submodularity concerns the extension of a submodular set-function discussed at the end of the last chapter.

4.1. Proposition. Let f be any set-function defined on the subsets of a set S and let $\hat{f}$ be its extension to non-negative vectors. Then $\hat{f}$ is convex (concave) if and only if f is submodular (supermodular).

A trivial consequence of this statement is that $\hat{f}$ is linear if and only if f is modular. Conversely, the restriction of a linear set-function to 01-vectors is always a modular set-function. The restriction of a convex function to 01-vectors,

24. MRZ. 1983
1. Hauptexemplar

however, does not always yield a submodular set-function; consider e.g. $f(x_1, x_2, x_3) = \max\{x_1, x_2+x_3\}$.

This proposition can be used very well in the study of submodularity. As a first application, we shall use it to prove the first half of the following "sandwich theorem" due to A. Frank (1980).

4.2. Theorem. Let f be a submodular and g a supermodular set-function, both defined on the subsets of the same set S, and assume that $g \leq f$.

(a) Then there exists a modular set-function h on the subsets of S such that $g \leq h \leq f$.

(b) If f and g are integral valued then this separating set-function h can also be chosen integral valued.

Proof. (a) By proposition 4.1, $\hat{f}$ is convex and $\hat{g}$ is concave, and it follows from the construction of $\hat{f}$ and $\hat{g}$ that $\hat{g} \leq \hat{f}$. By an elementary result in convexity, there exists a linear function $\hat{h}$ such that $\hat{g} \leq \hat{h} \leq \hat{f}$ for all non-negative vectors. The restriction of $\hat{h}$ to 01-vectors yields a modular set-function separating f and g.

(b) This statement is considerably deeper and more interesting from the combinatorial point of view. It is in fact equivalent to Corollary 3.4. It can also be proved by mimicking (mutatis mutandis) a standard proof of the Hahn-Banach Theorem.

Perhaps the most significant application of Proposition 4.1 is the minimization of a submodular set-function. Let f be a submodular set-function defined on the subsets of S. We want to find the set $X \subseteq S$ for which $f(X)$ is minimum. It turns out that this is a very general and important problem in combinatorial optimization and very many combinatorial problems can be reduced to it – we shall see some in section 6.

Speaking about algorithms concerning submodular set-functions, we have to specify first how the set-function is given. Most problems which arise would become trivial if the set-function involved were given by listing all of its 2^n values. Therefore, we shall assume that our set-function is given by an oracle, i.e. subroutine which evaluates the function at any subset. One call of this subroutine counts as one step only, and the subroutine cannot be "broken open", i.e. the algorithm must work for any subroutine whose output is a real number for every input (a subset of S), provided these outputs form a submodular set-function, regardless of the way of calculating this value.

The key to the minimization of a submodular set-function is the following lemma.

4.3. Lemma. Let f be a set-function defined on the subsets of S such that $f(\emptyset) = 0$. Then

$$\min\{f(X) : X \subseteq S\} = \min\{\hat{f}(x) : x \in [0, 1]^S\}.$$

The proof of this lemma is straightforward. So instead of minimizing f over the subsets of S, it suffices to minimize $\hat{f}$ over the cube. Now this is made possible by the following two nice properties of the function $\hat{f}$:

24. MRZ. 1983
1. Hauptkorrektur

a) for every rational vector $x \in [0, 1]^S$, $f(x)$ can be evaluated in polynomial time;

b) if f is submodular then f is convex.

Thus we can apply an algorithm due to Judin and Nemirovskii (1976) and find the minimum of f in polynomial time. Since the hypothesis that $f(\emptyset)=0$ is clearly irrelevant (we can replace, if necessary, f by $f-f(\emptyset)$), we have proved the following result (Grötschel, Lovász and Schrijver 1981):

4.4. Theorem. Let f be a submodular set-function defined on the subsets of S. Then the subset of S minimizing f can be found in polynomial time.

This theorem implies the polynomial time solvability of many problems in combinatorial optimization. Let us mention some.

4.5. Corollary. Let f be a submodular set-function such that $f(\emptyset)=0$. Then $x \in P_f$ can be checked in polynomial time for every rational vector x.

4.6. Corollary. Let f and g be submodular set-functions defined on the subsets of S. Then $(f * g)(X)$ can be evaluated for every $X \subseteq S$ in polynomial time. In particular, $f_{mon}(X)$ can be evaluated in polynomial time.

Finally, the following results concerning the Dilworth truncation was shown by Grötschel, Lovász and Schrijver (1981).

4.7. Theorem. Let f be a submodular set-function defined on the subsets of S. Then $f_*(X)$ can be evaluated for every $X \subseteq S$ in polynomial time.

Thus we have seen that the operations introduced in section 2, when applied to submodular set-functions given by an oracle, yield polynomially computable set-functions. In particular, these operations preserve the polynomial-time computability of the set-functions involved.

5. Submodularity and concavity

Let us hear now some arguments of the advocate for concavity. He will start with pointing out the analogy between propositions 1.1 and 2.2 and some properties of concave functions. A closer connection is represented by the following proposition:

5.1. Proposition. Let f be a real-valued function defined on non-negative integers. Then $f(|X|)$ is submodular on the subsets of an arbitrary set if and only if f is concave.

We have seen that the problem of minimizing a submodular set-function is an important one, whose solution implies the solution of many very general combinatorial optimization problems. Let us show now that the problem of maximizing a submodular set-function is also very important.

Let H be a hypergraph all whose edges are r-tuples of the set V, and let S be the set of its edges. It is a very important combinatorial problem to determine

24. MRZ. 1983
1. Hauskorrektur

the maximum number of disjoint edges of H. This can be formulated as a submodular function maximization problem as follows. For every $X \subseteq S$, let $f(X)$ denote the number of points of V covered by the edges in X. As remarked in example 1.4, this set-function is submodular. Therefore so is the set-function $f(X)-(k-1)|X|$. It is easy to verify that the maximum of this set-function is equal to the maximum number of disjoint edges of H.

One could show by a similar construction that the problem of finding maximum common independent set of k matroids is also a special case of submodular function minimization.

But these remarks are bad news rather than good. Since several NP-complete problems can be reduced to the problem of minimizing a submodular set-function, there is no hope to find a polynomial-time algorithm solving this problem (the problem of maximizing a submodular set-function is not in NP, so we cannot conclude that it is NP-complete, but it is certainly NP-hard). Further, it follows from the results of Hausmann-Korte (1978) and Lovász (1981) (independently of the $P \neq NP$ hypothesis) that every algorithm finding the minimum of submodular functions (even of 1-smooth submodular functions) takes exponential time in the worst case.

But this last mentioned special case of 1-smooth submodular functions is not as bad as it seems. No special choice of the submodular function is known to lead to an NP-complete problem, and quite a few special choices are solvable in polynomial time. This problem is in fact equivalent to the matroid matching (matroid parity, or matchoid) problem, which can be stated as follows.

Let f be a polymatroid function such that $f(\{x\})=2$ for every $x \in S$. We shall call f a *2-polymatroid function*. It is clear that $f(X) \leq 2|X|$ holds for every $X \subseteq S$. The subset X is called a *matching* if $f(X)=2|X|$. The *matroid matching problem* asks for the maximum size of a matching for a given 2-polymatroid function f.

It is easy to see that the maximum size of a matching is equal to the maximum of the submodular set-function $f(X)-|X|$.

Now the good news is that a maximum matching can be found in polynomial time provided the 2-polymatroid is given

(a) as a set of subspaces of a linear space (Example 1.8; Lovász 1981), or
(b) as a set of pairs of points of a transversal matroid (Lawler and Po Tong 1982).

(One may wonder why to state (b) as transversal matroids are known to be linear and therefore the 2-polymatroids (b) can also be represented in the form of (a). No efficient way is known, however, to construct such a representation.)

The problem of maximizing a t-smooth submodular function is NP-hard for every $t \geq 2$. But submodularity and t-smoothness do help in finding reasonably good approximations of the optimum; see Fisher-Nemhauser-Wolsey (1978).

Spitzkolumne

24. MRZ. 1983
1. Hauskorrektur

6. Submodular objective functions and constraints

DT2:8227X3XX50.0S

The set-function which occurs by far the most often in combinatorics is the cardinality function, which is a modular and 1-smooth set-function. In weighted and polyhedral versions of combinatorial optimization problems we deal with more general modular functions. In many proof techniques, however, one can further generalize and use only sub- (or super-) modularity of the set-functions in question. It turns out that this generalization is not all all l'art pourt l'art, but it yields beautiful results which contain many deep combinatorial min-max results as special cases, while in their statements and proofs, as well as in the algorithms that go with them, they preserve the simplicity and heuristic value of the original purely combinatorial results.

As an early example of this approach, let us consider a version of the Matroid Intersection Theorem, due to Aigner and Dowling (1971). Let G be a graph and let $(V(G), \mathscr{M})$ be a matroid defined on its nodes. Let us say that a matching of G is a *matroid matching* if the set of nodes of G it meets is independent in the matroid. Then the following generalization of the König Theorem can be proved.

6.1. Theorem. Let G be a bipartite graph and $(V(G), \mathscr{M})$ a matroid such that the color-classes of G are separators of $(V(G), \mathscr{M})$. Then the maximum size of a matroid matching is equal to the minimum rank of a point-cover of G.

Next we discuss some more recent results on graphs featuring submodular constraints. Edmonds and Giles (1977) discuss *submodular flows*. Let G be a directed graph, $\mathscr{H} \subseteq 2^{V(G)}$ a crossing family and f a submodular function defined on $\mathscr{H}$. Let, further, two functions c, d: $E(G) \to \mathbb{R}$ be given such that $c(e) \geqq d(e)$ for all $e \in E(G)$. A *submodular flow* defined on G is an assignement of real numbers $x(e)$ to the lines e such that

$$d(e) \leqq x(e) \leqq c(e) \quad \text{for all } e \in E(G), \tag{8}$$

$$x(\nabla^+(T)) - x(\nabla^-(T)) \leqq f(T) \quad \text{for all } T \in \mathscr{H}. \tag{9}$$

(Here $\nabla^+(T)$ and $\nabla^-(T)$ denote the sets of lines leaving and entering T, respectively.) The result of Edmonds and Giles states:

6.2. Theorem. If f, c and d are integral valued functions then the system (8)-(9) is total dual integral. Consequently, for any weighting of the lines wiht integers there exists a maximum weight submodular flow wiht integral values.

Perhaps the most important special case of this result is the Lucchesi-Younger Theorem on diconnecting sets. Let $\mathscr{H}$ consist of those non-empty proper subsets T of $V(G)$ which determine a directed cut, i.e. which have $\nabla^-(T) = \emptyset$. It is easy to see that this is a crossing family. Further, the function $\nabla^+(X) - 1$ is submodular on $\mathscr{H}$. Take $d = 0$ and $c = 1$ for every line. Then any submodular flow with integral values may be viewed as a subgraph, and (9) says that the complement of this subgraph meets every directed cut. So the integral submodular flows are precisely the complements of diconnecting sets. Taking all weights equal to 1, and applying linear programming duality to-

24. MRZ. 1983
1. Hauskorrektur

gether with the primal and dual integrality implied by the Edmonds-Giles theorem, we obtain the Lucchesi-Younger Theorem (1978):

6.3. Corollary. The minimum size of a diconnecting set in a digraph is equal to the maximum number of mutually disjoint directed cuts.

Lawler and Martell (1980) introduce a similar problem. While in the problem of Edmonds and Giles classical network flows are generalized by retaining capacity constraints but generalizing the "Kirchhoff" laws, Lawler and Martell retain the Kirchhoff laws but replace capacity constraints by submodular ones. To be more precise, let G be a directed graph, s and t two specified nodes of G, and let, for each node $v \in V(G)$, two submodular functions α_v, β_v be given: α_v on the subsets of $\nabla^+(v)$ and β_v on the subsets of $\nabla^-(v)$. The functions α_v yield a submodular set-function on the set $E(G)$ by taking their direct sum, and let β be similarly the direct sum of the functions β_v. Then a *polymatroidal network flow* is defined as an st-flow in the usual sense, which in addition satisfies the following conditions:

$$x(T) \leq \min\{\alpha(T), \beta(T)\} \quad \text{for all } T \subseteq E(G)$$

(it would suffice to reqire this for the subsets of in- and out-stars of nodes). Let, for each st-cut C, the capacity of C be defined as $(\alpha * \beta)(C)$.

With this definition, the following generalization of the Max-flow-min-cut Theorem can be proved by an appropriate refinement of augmenting path techniques:

6.4. Theorem. The maximum value of a polymatroidal network flow is equali to the minimum capacity of an st-cut.

A further closely related class of results is due to Frank (1979, 1980). We only state one of these. Let G be a graph and let f be a supermodular set-function defined on the subsets of $V(G)$. Let us say that an orientation $\vec{G}$ of G satiesfies demand f if

$$|\nabla^+_{\vec G}(T)| \geq f(T)$$

$\nabla^+_{\vec G}$

holds true for every subset $T \subseteq V(G)$, $T \neq \emptyset$, $V(G)$. Thus an orientation which satiesfies demand 1 is just strongly connected.

6.5. Theorem. The graph G has an orientation which satisfies demand f if and only if

$$\sum_{i=1}^{k} f(T_i) \leq |E(G)| \quad \text{and} \quad \sum_{i=1}^{k} f(S - T_i) \leq |E(G)|$$

holds for every partition $\{T_1, \ldots, T_k\}$ of $V(G)$.

Note that the first condition says that $f^*(V(G)) \leq |E(G)|$, and the second is obtained from the first by complementation in the variable.

We conclude with a different version of the same idea. Let us return to Theorem 6.1. What happens if we have a non-bipartite graph G or a bipartite graph whose color-classes are not separators of the underlying matroid? Already for ordinary graphs, the problems of maximum matching and minimum point-

24. MRZ. 1983
1. Mathematik

cover are different in the non-bipartite case, so we have to discuss them separately.

The problem of finding a maximum matroid matching is equivalent to the matroid matching problem as formulated in section 5. Thus, this problem is well-solved e. g. if the matroid is given as the columns of a matrix.

The problem of finding the minimum rank of a point-cover is NP-complete already in the case when the underlying matroid is free. So no complete solution of this problem can be expected. It is, however, possible to generalize various results on the point-covering number to this matroidal point-covering number. Applying these matroidal results to situations where the underlying matroids are not free (e. g. they are transversal), further graph-theoretic results can be proved. In some cases this generalization to matroids is the key to the proof. This is the case with a finite basis theorem for τ-critical graphs (see Lovász 1977).

7. Concluding remarks

In this last section we mention some open problems concerning submodularity, without aiming at a complete list.

1. The algorithm presented in section 4 to minimize a submodular function (which is probably the key problem in the algorithmic theory of submodularity) is not at all satisfactory. It is polynomial in running time, but it involves the ellipsoid method which renders it more or less useless in practical applications. It is an outstanding open problem to find an algorithm to minimize a submodular set-function in polynomial time, which would operate by combinatorial means and which would be also practical.

2. The results discussed in section 6 indicate that a more thorough study of combinatorial optimization problems with submodular objective functions and constraints is in progress. Do further standard combinatorial optimization problems (matching, chromatic number, feedback etc.) have generalizations in this spirit which lead to interesting results?

3. The submodular functions defined on a given set form a convex cone. Some properties of this cone have been studied (Nguyen 1978), but its description is by far not satisfactory.

4. The functions δ_w of example 1.3 also form a convex cone. As opposed to the cone of submodular functions, for this cone the extreme rays are easy to find (they are the functions δ_w where w is 0 on all lines except one), so the question is to find a system of linear inequalities describing this cone. More modestly, we may ask for interesting valid inequalities. Submodularity is one of them, but there are others, e. g. the following family of "generalized submodular inequalities":

$$\sum_{i=1}^{k} f(X_i) \geq \sum f(X_1^{\varepsilon_1} \cap \ldots \cap X_k^{\varepsilon_k}),$$

24. MRZ. 1983

where $X^1 = X$, $X^{-1} = S - X$, and the summation on the right hand side extends to all ± 1-sequences $(e_1, \ldots, e_k)$ with $e_1 \ldots e_k = 1$.

5. The Dilworth truncation of a submodular function may be viewed as the most economical way to make $f(\emptyset)$ non-negative (at the expense of decreasing the function value on other sets). Is there a similar construction to make $f(\emptyset)$ and $f(\{x\})$ non-negative for every $x \in S$? More generally, which systems of subsets of S have the property that there is a canonical way to increase the value of a submodular function on these sets at the expense of decreasing it on the others?

6. Let f be a polymatroid function defined on the subsets of a set S and let $x, y \in S$. When does f have an extension f' to the set $S \cup \{z\}$ such that $f'(z) = 1$, $f'(\{x, z\}) = f(\{x\})$ and $f'(\{y, z\}) = f(\{y\})$? Geometrically, this would mean to place a point on the intersection of the subspaces x and y. This is probably a very difficult problem in general, but as the Dilworth truncation shows, it can be solved if e.g. x is a "general hyperplane". Can less restrictive conditions be found under which this extension problem can be solved?

7. One may also raise the question whether any reasonable class of set-functions more general than submodular ones admits a similarly deep theory. There are some exmaples which direct in the direction of a positive answer. The Polymatroid Intersection Theorem shows that the minimum of two submodular set-functions, even though not submodular in general, gives rise to a polyhedron which still has nice properties (e.g. integral vertices). Results of Hoffman and Gröflin (1981) show that an analogue of the Matroid Partition Theorem is true for the rank function of the intersection of two matroids, which is not submodular but the convolution of two submodular functions. Does there exist a property similar to, but more general than, submodularity which plays role in these examples?

References

M. Aigner, T. A. Dowling (1971), Matching theory for combinatorial geometrics, *Trans. Amer. Math. Soc.* 158, 231–245.

L. Choquet (1955), Theory of capacities, *Ann. Inst. Fournier Grenoble* 5, 131–295.

H. Crapo (1965), Single-element extensions of matroids, *J. Res. Nat. Bur. Stand.* 69 B, 55–65.

H. Crapo (1979), Structural rigidity, *Structural Topology* 1, 26–45.

R. P. Dilworth (1944), Dependence relations in a semimodular lattice, *Duke Math. J.* 11, 575–586.

F. D. J. Dunstan (1976), Matroids and submodular functions, *Quart. J. Math. Oxford* 27, 339–348.

J. Edmonds (1970), Submodular functions, matroids, and certain polyhedra, in: *Combinatorial Structures and their Applications* (eds. R. Guy, H. Hanani, N. Sauer, J. Schönheim) Gordon and Breach, 69–87.

J. Edmonds, D. R. Fulkerson (1965), Transversals and matroid partition, *J. Res. Nat. Bur. Stand.* 69 B, 147–153.

DT2:8227X3XX51.0

24. MRZ. 1983

J. Edmonds, R. Giles (1977), A min-max relation on submodular functions on graphs, *Annals of Discrete Math.* 1, 185–204.

M. L. Fisher, G. L. Nemhauser, L. A. Wolsey (1978), Analysis of approximations for maximizing a submodular setfunction II, *Math. Prog. Study* 8, 73–87.

A. Frank (1980), An algorithm for submodular functions on graphs, *Annals of Discrete Math.*

A. Frank (1980), On the orientation of graphs, *J. Comb. Theory* B28, 251–261.

A. Frank (1979), Kernel systems of directed graphs, *Acta Sci. Math. Univ. Szeged* 41, 63–76.

H. Gröflin, A. J. Hoffman (1981), On matroid intersections, *Combinatorica* 1, 43–47.

M. Grötschel, L. Lovász, A. Schrijver (1981), The ellipsoid method and its consequences in combinatorial optimization, *Combinatorica* 1, 169–197.

P. M. Jensen, B. Korte (1978), Complexity of matroid property algorithms, Rept. No. 7124-OR, Inst. Ökon. Oper. Res. Univ. Bonn.

T. Helgason (1974), Aspects of the theory of hypermatroids, in: *Hypergraph Seminar* (eds. C. Berge, D. K. Ray-Chaudhuri), Lecture Notes in Math. 411, Springer, 191–214.

D. B. Iudin, A. S. Nemirovskii (1976), Informational complexity and effective methods of solution for convex extremal problems, *Ekon. i Mat. Met.* 12, 357–369; *Matekon* 13 (3), 24–45.

E. L. Lawler, C. U. Martell (1980), Computing maximal "polymatroidal" network flows, Res. Rep. UCB/ERL M80/52.

E. L. Lawler, Po Tong (1982), Lecture at the Conference on Combinatorics and Graph Theory, Univ. of Waterloo.

L. Lovász (1977), Flats in matroids and geometric graphs, in: *Combinatorial Surveys*, (ed. P. Cameron), Acad. Press, 45–86.

L. Lovász (1981), The matroid matching problem, in: *Algebraic Methods in Graph Theory* (eds. L. Lovász, V. T. Sós), North-Holland, 495–517.

L. Lovász, Y. Yemini (1982), On generic rigidity in the plane, SIAM J. on Alg. Discr. Meth. 3, 91–99.

C. J. H. McDiarmid (1975), Rado's Theorem for polymatroids, Math. Proc. Cambridge Phil. Soc. 78, 263–281.

J. H. Mason (1977), Matroids as the study of geometrical configurations, in: *Higher Combinatorics* (ed. M. Aigner) Reidel, 133–176.

J. H. Mason (1981), Glueing matroids together: a study of Dilworth truncations and matroid analogus of exterior and symmetric powers, in: *Algebraic Methods in Graph Theory* (eds. L. Lovász, V. T. Sós), North-Holland, 519–561.

H. Q. Nguyen (1978), Semimodular functions and combinatorial geometries, *Trans. AMS* 238, 355–383.

O. Ore (1955), Graphs and matching theorems, *Duke Math. J.* 22, 625–639.

D. Welsh (1976), *Matroid Theory*, Academic Press.

24. MRZ. 1983

12