No. 198 (83-21)

ON THE SUBDIFFERENTIAL OF A SUBMODULAR FUNCTION

by

Satoru FUJISHIGE

November 1983

Title: On the Subdifferential of a Submodular Function *

Author: Satoru FUJISHIGE

Affiliation: Institute of Socio-Economic Planning,

University of Tsukuba, Sakura, Ibaraki 305, Japan

Abstract: The author recently introduced a concept of a subdifferential of a submodular function defined on a distributive lattice. Each subdifferential is an unbounded polyhedron. In the present paper we determine the set of all the extreme points and rays of each subdifferential and show the relationship between subdifferentials of a submodular function and subdifferentials, in an ordinary sense of convex analysis, of Lovász's extension of the submodular function. Furthermore, for a modular function on a distributive lattice we give an algorithm for determining which subdifferential contains a given vector and finding a nonnegative linear combination of extreme vectors of the subdifferential which expresses the given vector minus the unique extreme point of the subdifferential.

Keywords: Submodular functions, subdifferentials, polyhedra, extreme points, extreme rays, Lovász's extensions of set functions.

Abbreviated title: Subdifferential of a submodular function.

---

*) The present research was carried out when the author was on leave at Institut fur Ökonometrie und Operations Research, Universität Bonn, and was supported by the Alexander von Humboldt Fellowship (1982/83), West Germany.

1. Introduction

The author [6] introduced a concept of a subdifferential $\partial f(X)$ $(X \in \mathcal{D})$ of a submodular function $f: \mathcal{D} \rightarrow R$ from a distributive lattice $\mathcal{D} \subseteq 2^E$ to the set R of reals and developed a theory of submodular programs. (A precise definition of a subdifferential will be given later.) Each subdiferential $\partial f(X)$ $(X \in \mathcal{D})$ is a polyhedron which plays a fundamental role in minimization of the submodular function f.

In the present paper we determine, for each $X \in \mathcal{D}$, the set of all the extreme points and rays of the subdifferential $\partial f(X)$ which generates $\partial f(X)$ and show the relationship between subdifferentials of a submodular function and subdifferentials, in an ordinary sense of convex analysis, of Lovász's extension of the submodular function. Furthermore, for a modular function on a distributive lattice we give an algorithm for determining which subdifferential contains a given vector and finding a nonnegative linear combination of extreme vectors of the subdifferential which expresses the given vector minus the unique extreme point of the subdifferential.

2. Definitions and Preliminaries

Let E be a finite set and denote by 2^E the collection of all the subsets of E. Throughout the present paper we consider a distributive lattice $\mathcal{D} \subseteq 2^E$ formed by all the ideals of a partially ordered set $\mathcal{P} = (E, \preceq)$ with set union and intersection as the lattice operations, where $X \subseteq E$ is an ideal of $\mathcal{P} = (E, \preceq)$ if $e \preceq e' \in X$ implies $e \in X$ for all $e, e' \in E$. Note that by definition $\emptyset, E \in \mathcal{D}$. Also let $f\colon \mathcal{D} \to R$ be a submodular function on the distributive lattice $\mathcal{D}$, i.e., for every pair of $X, Y \in \mathcal{D}$

$$f(X) + f(Y) \geqq f(X \cup Y) + f(X \cap Y). \tag{2.1}$$

(If (2.1) holds with equality for every pair of $X, Y \in \mathcal{D}$, f is called a modular function. A function $g\colon \mathcal{D} \to R$ is called a supermodular function if $-g$ is a submodular function.) A subdifferential $\partial f(A)$ of a submodular function $f\colon \mathcal{D} \to R$ at $A \in \mathcal{D}$ is a polyhedron given by

$$\partial f(A) = \{x \mid x \in R^E, \forall X \in \mathcal{D}\colon x(X) - x(A) \leqq f(X) - f(A)\} \tag{2.2}$$

(see [6]), where R^E is the set of all the real-valued functions (or vectors) from E to R and for any $x \in R^E$ and $X \in \mathcal{D}$

$$x(X) = \sum_{e \in X} x(e). \tag{2.3}$$

Each $x \in \partial f(A)$ $(A \in \mathcal{D})$ is called a subgradient of f at A. Note that R^E is divided into $|\mathcal{D}|$ nonempty parts $\partial f(A)$ $(A \in \mathcal{D})$, where $|\cdot|$ denotes the cardinality, and for each distinct $A, B \in \mathcal{D}$ $\partial f(A)$ and $\partial f(B)$ may have common faces but do not have any common interior point in R^E.

The concept of a subdifferential of a submodular function is fundamental in the problem of minimizing submodular functions. It trivially follows from definition (2.2) that $A \in \mathcal{D}$ minimizes a submodular function

$f\colon \mathcal{D} \to R$ if and only if $\mathbb{0} \in \partial f(A)$, where $\mathbb{0}$ is the zero vector in R^E, as is an ordinary subdifferential of a convex function [13].

The system of inequalities appearing in (2.2) contains redundant inequalities. The following lemma is fundamental and deletes some of the redundant inequalities in (2.2).

Lemma 2.1: For each $A \in \mathcal{D}$ the subdifferential $\partial f(A)$ is also given by

$$\partial f(A) = \{x \mid x \in R^E,\ \forall X \in [\emptyset, A]_{\mathcal{D}} \cup [A, E]_{\mathcal{D}}\colon x(X) - x(A) \leqq f(X) - f(A)\}, \tag{2.4}$$

where $[\emptyset, A]_{\mathcal{D}}$ and $[A, E]_{\mathcal{D}}$ are intervals in $\mathcal{D}$ defined by

$$[\emptyset, A]_{\mathcal{D}} = \{X \mid X \in \mathcal{D},\ X \subseteq A\}, \tag{2.5}$$

$$[A, E]_{\mathcal{D}} = \{X \mid X \in \mathcal{D},\ A \subseteq X \subseteq E\}. \tag{2.6}$$

The proof of Lemma 2.1 is immediate and is also given in [6].

Corollary 2.2: For a submodular function $f\colon \mathcal{D} \to R$, $A \in \mathcal{D}$ minimizes f on $\mathcal{D}$ if and only if A minimizes f on $[\emptyset, A]_{\mathcal{D}} \cup [A, E]_{\mathcal{D}}$.

Consider a submodular function $f\colon \mathcal{D} \to R$ and a supermodular function $g\colon \mathcal{D} \to R$ with $f(\emptyset) = g(\emptyset) = 0$. The pair $(\mathcal{D}, f)$ is called a submodular system and the pair $(\mathcal{D}, g)$ a supermodular system (see e.g. [6]). A polyhedron

$$P(f) = \{x \mid x \in R^E,\ \forall X \in \mathcal{D}\colon x(X) \leqq f(X)\} \tag{2.7}$$

is called a submodular polyhedron associated with $(\mathcal{D}, f)$, and a polyhedron

$$B(f) = \{x \mid x \in P(f),\ x(E) = f(E)\} \tag{2.8}$$

is called a base polyhedron associated with $(\mathcal{D}, f)$.

Similarly, a polyhedron

$$P(g) = \{x \mid x \in R^E, \forall X \in \mathcal{D}: x(X) \geqq g(X)\} \tag{2.9}$$

is called a <u>supermodular polyhedron</u> and

$$B(g) = \{x \mid x \in P(g),\ x(E) = g(E)\} \tag{2.10}$$

a <u>base polyhedron</u> associated with $(\mathcal{D}, g)$. Define the dual of $\mathcal{D}$ by

$$\overline{\mathcal{D}} = \{E - X \mid X \in \mathcal{D}\} \tag{2.11}$$

and a supermodular function $f^{\#}: \overline{\mathcal{D}} \to R$ by

$$f^{\#}(E - X) = f(E) - f(X) \qquad (X \in \mathcal{D}), \tag{2.12}$$

where $f^{\#}$ is called the <u>dual supermodular function</u> of f. Note that, when $f(\emptyset) = 0$,

$$\partial f(\emptyset) = P(f), \tag{2.13}$$

$$\partial f(E) = P(f^{\#}). \tag{2.14}$$

Furthermore, we have the following

<u>Lemma 2.3</u> ([4], [15]): For a submodular function $f: \mathcal{D} \to R$ with $f(\emptyset) = 0$ we have

$$B(f) = B(f^{\#}). \tag{2.15}$$

Lemma 2.3 corresponds to the polymatroid duality [11]. From (2.13) - (2.15),

$$\partial f(\emptyset) \cap \partial f(E) = B(f) = B(f^{\#}) \tag{2.16}$$

which is a maximal common face of $\partial f(\emptyset)$ and $\partial f(E)$, and the polyhedra $\partial f(\emptyset)$, $\partial f(E)$ and $B(f)$ $(= B(f^{\#}))$ have the same set of extreme points. The extreme points of $B(f)$ are characterized as follows.

Lemma 2.4 [7]: A vector $x \in R^E$ is an extreme point of $B(f)$ if and only if there exists a maximal chain

$$\mathcal{C}: \emptyset = S_0 \subsetneqq S_1 \subsetneqq \cdots \subsetneqq S_n = E \tag{2.17}$$

in $\mathcal{D}$ such that

$$x(S_i - S_{i-1}) = f(S_i) - f(S_{i-1}) \qquad (i=1,2,\ldots,n). \tag{2.18}$$

Here, note that from the assumption on $\mathcal{D}$ we have $|S_i - S_{i-1}| = 1$ $(i=1,2,\ldots,n)$ in (2.17) and (2.18). Lemma 2.4 is a generalization of a result of Edmonds [1] for polymatroids (also see [10]).

3. Determination of Extreme Points and Rays of a Subdifferential

First, we give a characterization of the extreme points of the subdifferential $\partial f(A)$ for each $A \in \mathcal{D}$.

Theorem 3.1: For each $A \in \mathcal{D}$, $x \in R^E$ is an extreme point of $\partial f(A)$ if and only if there exists a maximal chain

$$\mathcal{C}: \emptyset = S_0 \subsetneqq S_1 \subsetneqq \cdots \subsetneqq S_n = E \tag{3.1}$$

in $\mathcal{D}$, including A in it, such that

$$x(S_i - S_{i-1}) = f(S_i) - f(S_{i-1}) \qquad (i=1,2,\ldots,n). \tag{3.2}$$

(Proof) We can assume $f(\emptyset) = 0$. From (2.13) we have $\partial f(\emptyset) = P(f)$ and $P(f)$ and $B(f)$ have the same set of extreme points. Therefore, the present theorem for the case of $A = \emptyset$ follows from Lemma 2.4. Similarly, the present theorem for the case of $A = E$ follows from Lemma 2.4, since $\partial f(E) = P(f^{\#})$ from (2.14), $P(f^{\#})$ and $B(f^{\#})$ have the same set of extreme points, and $B(f^{\#}) = B(f)$ from Lemma 2.3. Furthermore, note that the system of inequalities in (2.4) is in a separable form as

$$\forall\, X \in [\emptyset,A]_{\mathcal{D}}:\ x(X) - x(A) \leqq f(X) - f(A), \tag{3.3}$$

$$\forall\, X \in [A,E]_{\mathcal{D}}:\ x(X) - x(A) \leqq f(X) - f(A). \tag{3.4}$$

Let f' be a restriction of f to $[\emptyset,A]_{\mathcal{D}}$ and f'' be a contraction of f to $[A,E]_{\mathcal{D}}/A$ defined by

$$[A,E]_{\mathcal{D}}/A = \{X - A \mid X \in [A,E]_{\mathcal{D}}\}, \tag{3.5}$$

$$f''(X - A) = f(X) - f(A) \qquad (X \in [A,E]_{\mathcal{D}}). \tag{3.6}$$

Then it follows from (3.3) and (3.4) that $\partial f(A)$ is the direct product of $\partial f'(A)$ and $\partial f''(\emptyset)$ and thus that the extreme points of $\partial f(A)$ are given by the direct product of extreme points of $\partial f'(A)$ and those of

$\partial f''(\emptyset)$. Since A is the maximum element of the domain $[\emptyset, A]_{\mathcal{D}}$ of f' and $\emptyset$ is the minimum element of the domain $[A,E]_{\mathcal{D}}/A$ of f'', the present theorem for $\emptyset \subsetneqq A \subsetneqq E$ follows from that for $A = E$ and $A = \emptyset$ for f with the domain $\mathcal{D}$. Q.E.D.

Since for each $A \in \mathcal{D}$ the subdifferential $\partial f(A)$ is a polyhedron having at least one extreme point due to Theorem 3.1, $\partial f(A)$ $(A \in \mathcal{D})$ is expressed as

$$\partial f(A) = Q_f(A) + C_f(A), \tag{3.7}$$

where $Q_f(A)$ is a convex hull of all the extreme points of $\partial f(A)$ and $C_f(A)$ is the convex cone given by

$$\begin{aligned} C_f(A) &= \{x \mid x \in R^E, \forall X \in \mathcal{D}: x(X) - x(A) \leqq 0\} \\ &= \{x \mid x \in R^E, \forall X \in [\emptyset, A]_{\mathcal{D}} \cup [A,E]_{\mathcal{D}}: x(X) - x(A) \leqq 0\} \end{aligned} \tag{3.8}$$

called the recession cone of $\partial f(A)$.

Next, we give the extreme rays of $\partial f(A)$. (Here, we mean by an extreme ray of $\partial f(A)$ an extreme ray of the recession cone of $\partial f(A)$. An extreme ray of a cone is a class of projectively equivalent extreme vectors of the cone and is represented by an extreme vector in the class.)

Let $G(\mathcal{P}) = (E, B^*(\mathcal{P}))$ be a directed graph with a vertex set E and an arc set $B^*(\mathcal{P})$ defined as follows. An ordered pair (e,e') with $e, e' \in E$ is an arc in $B^*(\mathcal{P})$ if and only if $e' \precneqq e$ and there exists no $e'' \in E$ such that $e' \precneqq e'' \precneqq e$. That is to say, $G(\mathcal{P})$ is the Hasse diagram which represents the partially ordered set $\mathcal{P} = (E, \preceq)$. Let us denote by E^+ and E^-, respectively, the set of all the maximal and the minimal elements of $\mathcal{P} = (E, \preceq)$. Note that $E^+ \cap E^-$ may be nonempty.

For each $A \in \mathcal{D}$ also denote by $\Delta^{-}(A)$ the set of all the arcs (e,e') $\in B^{*}(\mathcal{P})$ such that $e \notin A$ and $e' \in A$. Let us define vectors $\xi(p^{+})$ $(p^{+} \in E^{+})$, $\eta(p^{-})$ $(p^{-} \in E^{-})$ and $\zeta(a)$ $(a \in B^{*}(\mathcal{P}))$ in R^{E} as follows.

$$\xi(p^{+})(e) = \begin{cases} -1 & (e = p^{+}) \\ 0 & (e \in E - \{p^{+}\}) \end{cases} \qquad (p^{+} \in E^{+}) \tag{3.9}$$

$$\eta(p^{-})(e) = \begin{cases} 1 & (e = p^{-}) \\ 0 & (e \in E - \{p^{-}\}) \end{cases} \qquad (p^{-} \in E^{-}) \tag{3.10}$$

$$\zeta(a)(e) = \begin{cases} 1 & (e = e') \\ -1 & (e = e'') \qquad (a = (e',e'') \in B^{*}(\mathcal{P})) \\ 0 & (e \in E - \{e',e''\}) \end{cases} \tag{3.11}$$

For each $A \in \mathcal{D}$, also define

$$ER(A) = \{\xi(p^{+}) \mid p^{+} \in E^{+} - A\} \cup \{\eta(p^{-}) \mid p^{-} \in E^{-} \cap A\}$$
$$\cup \{\zeta(a) \mid a \in B^{*}(\mathcal{P}) - \Delta^{-}(A)\}. \tag{3.12}$$

Then we have the following theorem.

<u>Theorem 3.2</u>: For the convex cone $C_f(A)$ appearing in the decomposition (3.7) of the subdifferential $\partial f(A)$, the set of all the extreme rays of $C_f(A)$ is given by $ER(A)$ in (3.12).

(Proof) Because of the expression (3.8) of the convex cone $C_f(A)$, we can easily see that

$$ER(A) \subseteqq C_f(A) \tag{3.13}$$

and that each vector in $ER(A)$ can not be expressed as a nonnegative linear combination of the vectors in $ER(A)$. Therefore, it is sufficient to prove that every vector in $C_f(A)$ can be expressed as a nonnegative linear combination of vectors in $ER(A)$. Let v be an arbitrary vector in $C_f(A)$. From (3.8),

$$v(A-X) \geqq 0 \qquad (A \supseteqq X \in \mathcal{D}), \tag{3.14}$$

$$v(X-A) \leqq 0 \qquad (A \subseteqq X \in \mathcal{D}). \tag{3.15}$$

Suppose that each arc of $B^*(\mathcal{P}) - \Delta^-(A)$ has an infinite upper capacity and a zero lower capacity and that each arc in $\Delta^-(A)$ has zero upper and lower capacities. Then it easily follows from (3.14), (3.15) and the feasibility theorem for network flows [3], [8] that there exists a nonnegative flow $\phi: B^*(\mathcal{P}) \to R_+$ in $G(\mathcal{P})$ with $\phi(a) = 0$ $(a \in \Delta^-(A))$, a nonpositive vector $x \in R_-^E$ with $x(e) = 0$ $(e \notin E^+ - A)$ and a nonnegative vector $y \in R_+^E$ with $y(e) = 0$ $(e \notin E^- \cap A)$ such that

$$v = \partial'\phi + x + y \tag{3.16}$$

where R_+ (R_-) is the set of all nonnegative (nonpositive) reals and $\partial'\phi$ is the boundary of ϕ defined by

$$\partial'\phi(e) = \sum_{(e,e') \in B^*(\mathcal{P})} \phi((e,e')) - \sum_{(e'',e) \in B^*(\mathcal{P})} \phi((e'',e)) \tag{3.17}$$

for each $e \in E$. This completes the proof. Q.E.D.

Here, the symbol ∂' for the boundary of a flow should not be confused with the symbol ∂ for a subdifferential. It should also be noted that if v satisfies $v(A) = 0$ in (3.14), then $y = \mathbb{0}$ and that if v satisfies $v(E-A) = 0$ in (3.15), then $x = \mathbb{0}$.

<u>Theorem 3.3</u> (Tomizawa): The set of all the extreme rays of the cone defined by

$$C_f' = \{x \mid x \in R^E,\ \forall X \in \mathcal{D}: x(X) \leqq 0,\ x(E) = 0\} \tag{3.18}$$

is given by

$$\{\zeta(a) \mid a \in B^*(P)\}. \tag{3.19}$$

(Proof) Note that $C_f' = C_f(\emptyset) \cap C_f(E)$ and that C_f' is the common face of $C_f(\emptyset)$ and $C_f(E)$. Therefore, the present theorem follows from Theorem 3.2. Q.E.D.

Theorem 3.3 has recently been communicated from Tomizawa [16]. It should be noted that Theorem 3.3 also easily follows from the fact that the cone C_f' given by (3.18) is the collection of all the boundaries $\partial'\phi$ of flows ϕ in $G(P)$, where each arc in $G(P)$ has an infinite upper capacity and a zero lower capacity (cf. [17]).

4. Lovász's Extensions of Submodular Functions

For a submodular function $f: \mathcal{D} \to R$ with $f(\emptyset) = 0$, define a function $\hat{f}: R^E \to R \cup \{+\infty\}$ by

$$\hat{f}(c) = \max\{(c,x) \mid x \in \partial f(\emptyset)\}, \tag{4.1}$$

where

$$(c,x) = \sum_{e \in E} c(e)x(e). \tag{4.2}$$

Here, $\hat{f}$ is the support function of $\partial f(\emptyset)$ and is a positively homogeneous convex function [13], [14]. Note that $\hat{f}(c) < +\infty$ if and only if

$$c \in C_f^*(\emptyset) \equiv \{c' \mid \forall x \in C_f(\emptyset): (c',x) \leqq 0\}, \tag{4.3}$$

where $C_f(\emptyset)$ is the recession cone of $\partial f(\emptyset)$ (see (3.8)) and $C_f^*(\emptyset)$ is its polar cone. Condition (4.3) is equivalent to the one that $c: E \to R$ is a nonnegative monotone nonincreasing function from $\mathcal{P} = (E, \preceq)$ to $(R, \leqq)$ (cf. [7]). Therefore, for each $c \in C_f^*(\emptyset)$ there uniquely exist a monotone increasing sequence

$$A_1 \subsetneqq A_2 \subsetneqq \cdots \subsetneqq A_k \tag{4.4}$$

of nonempty $A_i \in \mathcal{D}$ $(i=1,2,\ldots,k)$ and positive $\lambda_i \in R$ $(i=1,2,\ldots,k)$ such that

$$c = \sum_{i=1}^{k} \lambda_i \chi(A_i), \tag{4.5}$$

where $k \geqq 0$, the empty sum is equal to zero, and $\chi(A_i) \in R^E$ is the characteristic vector of A_i $(i=1,2,\ldots,k)$. Moreover,

$$\hat{f}(c) = \sum_{i=1}^{k} \lambda_i f(A_i), \tag{4.6}$$

since $\hat{f}(c)$ defined by (4.1) can be obtained by the so-called greedy algorithm for $c \in C_f^*(\emptyset)$ (cf. [7]). The expression (4.6) (for $\mathcal{D} = 2^E$) is introduced by L. Lovász's [10] and the construction of $\hat{f}$ through

(4.4) - (4.6) can be applied to any function defined on $\mathcal{D}$. We call such an extension $\hat{f}$ <u>Lovász's extension</u> of f.

Define

$$P(\mathcal{D}) = \text{the convex hull of } \chi(A)\ (A \in \mathcal{D}) \subseteqq C_f{}^*(\emptyset). \qquad (4.7)$$

<u>Lemma 4.1</u>: Let $f: \mathcal{D} \to R$ be a submodular function with $f(\emptyset) = 0$. Then

$$\min\{f(X) \mid X \in \mathcal{D}\} = \min\{\hat{f}(c) \mid c \in P(\mathcal{D})\}. \qquad (4.8)$$

(See [10] for $\mathcal{D} = 2^E$.)

Now, define $\hat{\hat{f}}: R^E \to R \cup \{+\infty\}$ by

$$\hat{\hat{f}}(c) = \begin{cases} \hat{f}(c) & (c \in P(\mathcal{D})), \\ +\infty & (c \in R^E - P(\mathcal{D})). \end{cases} \qquad (4.9)$$

<u>Theorem 4.2</u>: Let $f: \mathcal{D} \to R$ be a submodular function with $f(\emptyset) = 0$. Then, for each $A \in \mathcal{D}$,

$$\partial f(A) = \partial\hat{\hat{f}}(\chi(A)). \qquad (4.10)$$

Here, $\partial\hat{\hat{f}}(\chi(A))$ denotes the subdifferential of the convex function $\hat{\hat{f}}: R^E \to R \cup \{+\infty\}$ at $\chi(A)$ in an ordinary sense of convex analysis [13].

(Proof) The condition that $x \in \partial\hat{\hat{f}}(\chi(A))$ is equivalent to

$$(c - \chi(A), x) \leqq \hat{\hat{f}}(c) - \hat{\hat{f}}(\chi(A)) \qquad (4.11)$$

for all $c \in R^E$. Since $\hat{\hat{f}}(\chi(A)) = f(A)$, condition (4.11) is equivalent to

$$\begin{aligned} f(A) - x(A) &= \min\{\hat{\hat{f}}(c) - (c,x) \mid c \in R^E\} \\ &= \min\{\hat{f}(c) - (c,x) \mid c \in P(\mathcal{D})\} \\ &= \min\{f(X) - x(X) \mid X \in \mathcal{D}\}. \end{aligned} \qquad (4.12)$$

Here, the last equality follows from Lemma 4.1 with f replaced by f - x.

Condition (4.12) is equivalent to the condition that $x \in \partial f(A)$. Q.E.D.

For any nonzero vector $c \in P(\mathcal{D})$, suppose c is expressed as (4.5) with a chain (4.4). Since $c \in P(\mathcal{D})$,

$$\sum_{i=1}^{k} \lambda_i \leqq 1, \qquad \lambda_i > 0 \quad (i=1,2,\ldots,k). \tag{4.13}$$

If $\sum_{i=1}^{k} \lambda_i < 1$, then let us express c as

$$c = \sum_{i=0}^{k} \lambda_i \chi(A_i), \tag{4.14}$$

where $\lambda_0 = 1 - \sum_{i=1}^{k} \lambda_i > 0$ and $A_0 = \emptyset$. In such a way, for any $c \in P(\mathcal{D})$, c is uniquely expressed as a convex combination:

$$c = \sum_{i=1}^{\ell} \mu_i \chi(B_i) \tag{4.15}$$

with

$$B_1 \subsetneqq B_2 \subsetneqq \cdots \subsetneqq B_\ell, \qquad B_i \in \mathcal{D} \quad (i=1,2,\ldots,\ell), \tag{4.16}$$

$$\sum_{i=1}^{\ell} \mu_i = 1, \qquad \mu_i > 0 \quad (i=1,2,\ldots,\ell). \tag{4.17}$$

Here, B_1 may be empty.

<u>Theorem 4.3</u>: For any $c \in P(\mathcal{D})$,

$$\partial\hat{\hat{f}}(c) = \cap\{\partial f(B_i) \mid i=1,2,\ldots,\ell\}, \tag{4.18}$$

where B_i $(i=1,2,\ldots,\ell)$ are those in (4.15) - (4.17) which express c as (4.15).

(Proof) For $c \in P(\mathcal{D})$, $x \in \partial\hat{\hat{f}}(c)$ if and only if

$$\forall b \in P(\mathcal{D}): (b-c,x) \leqq \hat{f}(b) - \hat{f}(c). \tag{4.19}$$

Then from (4.15) - (4.17), (4.19) is equivalent to

$$\sum_{i=1}^{\ell} \mu_i\{f(B_i) - x(B_i)\} \leqq \min\{\hat{f}(b) - (b,x) \mid b \in P(\mathcal{D})\}$$

$$= \min\{f(X) - x(X) \mid X \in \mathcal{D}\} \qquad (4.20)$$

because of Lemma 4.1. Furthermore, from (4.17), (4.20) is equivalent to

$$f(B_i) - x(B_i) = \min\{f(X) - x(X) \mid X \in \mathcal{D}\} \qquad (4.21)$$

for $i = 1, 2, \ldots, \ell$, or

$$x \in \cap\{\partial f(B_i) \mid i=1,2,\ldots,\ell\}. \qquad (4.22)$$

Q.E.D.

For any maximal chain $\mathcal{C}$: $\emptyset = S_0 \subsetneq S_1 \subsetneq \cdots \subsetneq S_n = E$ in $\mathcal{D}$, denote by $P(\mathcal{C})$ the n-simplex with vertices $\chi(S_i)$ $(i=0,1,\ldots,n)$.

Lemma 4.4: The collection of $P(\mathcal{C})$'s for all maximal chains $\mathcal{C}$ in $\mathcal{D}$ is a simplicial subdivision of $P(\mathcal{D})$.

It follows from Lemma 4.4 that the union of $P(\mathcal{C})$'s for all maximal chains $\mathcal{C}$ in $\mathcal{D}$ containing a fixed $A \in \mathcal{D}$ is a neighborhood of $\chi(A)$ in $P(\mathcal{D})$. This implies Corollary 2.2.

Lemma 4.5: Let $f: \mathcal{D} \to R$ be a submodular function with $f(\emptyset) = 0$. For any maximal chain $\mathcal{C}$: $\emptyset = S_0 \subsetneq S_1 \subsetneq \cdots \subsetneq S_n = E$ in $\mathcal{D}$ and any interior point c of $P(\mathcal{C})$, $\hat{\hat{f}}$ has a unique subgradient x at c given by

$$x(S_i - S_{i-1}) = f(S_i) - f(S_{i-1}) \quad (i=1,2,\ldots,n). \qquad (4.23)$$

(Proof) The present lemma follows from Theorems 3.1, 3.2 and 4.3.

Q.E.D.

We defined in [6] the convex conjugate function $f^*: R^E \to R$ of a submodular function $f: \mathcal{D} \to R$ by

$$f^*(x) = \max\{x(X) - f(X) \mid X \in \mathcal{D}\}. \tag{4.24}$$

By Lemma 4.1,

$$f^*(x) = \max\{(c,x) - \hat{\hat{f}}(c) \mid c \in R^E\}. \tag{4.25}$$

Therefore, f^* and $\hat{\hat{f}}$ are the convex conjugate function of each other in an ordinary sense [13]. Consequently, the Fenchel-type min-max theorem for submodular and supermodular functions on distributive lattices [6] easily follows from Fenchel's duality theorem for ordinary convex and concave functions except for the integrality property (cf. [6]).

5. Modular Functions on Distributive Lattices

Given a submodular function $f\colon \mathcal{D} \to R$ and a vector $x \in R^E$, consider the problem: (1) to find $A \in \mathcal{D}$ such that $x \in \partial f(A)$ and then (2) to find an expression

$$x = x_1 + x_2 \tag{5.1}$$

such that x_1 is a convex combination of extreme points of $\partial f(A)$ and x_2 is a nonnegative linear combination of extreme vectors of $C_f(A)$. It does not seem to be easy to solve the problem for general f but in the special case when f is a modular function on $\mathcal{D}$ we can easily solve it based on the results in Section 3.

In this section we suppose that $f\colon \mathcal{D} \to R$ is a modular function (with $f(\emptyset) = 0$).

Lemma 5.1: For a modular function $f\colon \mathcal{D} \to R$ with $f(\emptyset) = 0$, there exists a unique vector $\nu \in R^E$ such that for every $X \in \mathcal{D}$

$$f(X) = \sum_{e \in X} \nu(e). \tag{5.2}$$

Note that for each $A \in \mathcal{D}$ the subdifferential $\partial f(A)$ has the unique extreme point ν appearing in Lemma 5.1. Therefore, it follows from (3.1) that the above problem is reduced to that of finding $A \in \mathcal{D}$ such that $x - \nu \in C_f(A)$ and of finding a nonnegative linear combination of $ER(A)$ given by (3.12) which expresses $x - \nu$.

The proof of Theorem 3.2 suggests a solution algorithm as follows.

An Algorithm

(I) Find a nonnegative flow $\phi: B^*(\mathcal{P}) \to R_+$ in $G(\mathcal{P})$ and nonnegative coefficients $\alpha(p^+)$ $(p^+ \in E^+)$ and $\beta(p^-)$ $(p^- \in E^-)$ such that

$$\sum_{p^+ \in E^+} \alpha(p^+)\xi(p^+) + \sum_{p^- \in E^-} \beta(p^-)\eta(p^-) + \partial'\phi = x - \nu. \quad (5.3)$$

(Here, $\partial'\phi$ is defined by (3.17), $\xi(p^+)$ by (3.9), $\eta(p^-)$ by (3.10) and for $p^+ = p^- \in E^+ \cap E^-$ we choose the values of $\alpha(p^+)$ and $\beta(p^-)$ such that $\alpha(p^+)\beta(p^-) = 0$.)

(II) Consider a network $\hat{N} = (\hat{G}(\mathcal{P}), \hat{c})$ with an underlying graph $\hat{G}(\mathcal{P}) = (E, \hat{B}(\mathcal{P}))$ and a capacity function $\hat{c}$ defined as follows. The arc set $\hat{B}(\mathcal{P})$ of $\hat{G}(\mathcal{P})$ is defined by

$$\hat{B}(\mathcal{P}) = B^*(\mathcal{P}) \cup \{(e,e') \mid (e',e) \in B^*(\mathcal{P})\} \quad (5.4)$$

and the capacity function $\hat{c}$ by

$$\hat{c}(a) = \begin{cases} \phi(a) & (a \in B^*(\mathcal{P})) \\ +\infty & (a = (e,e'),\ (e',e) \in B^*(\mathcal{P})). \end{cases} \quad (5.5)$$

Then find a maximal flow $\psi: \hat{B}(\mathcal{P}) \to R_+$ in $\hat{N}$ from the entrance vertex set $E^+ - E^-$ to the exit vertex set $E^- - E^+$ such that

$$0 \leqq \psi(a) \leqq \hat{c}(a) \qquad (a \in \hat{B}(\mathcal{P})), \quad (5.6)$$

$$\partial'\psi(e) = 0 \qquad (e \in E - (E^+ \cup E^-)), \quad (5.7)$$

$$\partial'\psi(p^+) \leqq \alpha(p^+) \qquad (p^+ \in E^+ - E^-), \quad (5.8)$$

$$-\partial'\psi(p^-) \leqq \beta(p^-) \qquad (p^- \in E^- - E^+). \quad (5.9)$$

(Here, the boundary operator ∂' is defined with respect to $\hat{G}(\mathcal{P})$.)

(III) Put

$$\phi((e,e')) \leftarrow \phi((e,e')) - \psi((e,e')) + \psi((e',e)) \quad ((e,e') \in B^*(\mathcal{P})), \quad (5.10)$$

$$\alpha(p^+) \leftarrow \alpha(p^+) - \partial'\psi(p^+) \qquad (p^+ \in E^+ - E^-), \tag{5.11}$$

$$\beta(p^-) \leftarrow \beta(p^-) + \partial'\psi(p^-) \qquad (p^- \in E^- - E^+). \tag{5.12}$$

Then find $A \in \mathcal{D}$ such that

(i) $\phi(a) = 0 \qquad (a \in \Delta^-(A))$, (5.13)

(ii) $\alpha(p^+) = 0 \qquad (p^+ \in E^+ \cap A)$, (5.14)

(iii) $\beta(p^-) = 0 \qquad (p^- \in E^- - A)$. (5.15)

For any $A \in \mathcal{D}$ satisfying (i) - (iii) we have $x \in \partial f(A)$ and x is expressed as

$$x = \nu + \sum_{p^+ \in E^+} \alpha(p^+)\xi(p^+) + \sum_{p^- \in E^-} \beta(p^-)\eta(p^-) + \sum_{a \in B^*(\mathcal{P})} \phi(a)\zeta(a). \tag{5.16}$$

Remark 5.1: Because of (5.13) - (5.15) $x - \nu$ in (5.16) is a nonnegative linear combination of $ER(A)$.

Remark 5.2: Step (I) can be carried out with a breadth-first type method (toward the roots) by considering a spanning branching T_1 directed from the set E^+ of roots (i.e., each vertex $e \in E - E^+$ has a unique arc (e',e) in T_1) and a spanning co-branching T_2 directed to the set E^- of roots (i.e., each vertex $e \in E - E^-$ has a unique arc (e,e') in T_2). T_1 and T_2 are used for adjusting the value of the boundary $\partial'\phi(e)$ for e with $x(e) - \nu(e) < 0$ and $x(e) - \nu(e) > 0$, respectively. Step (I) requires $O(|E| + |B^*(\mathcal{P})|)$ running time.

Remark 5.3: Step (II) is performed by ordinary maximum flow algorithms, which requires $O(|E|^3)$ running time or the less. Any minimum cut obtained by performing Step (II) gives a desired $A \in \mathcal{D}$ in Step (III).

Remark 5.4: When $x = \mathbb{0}$, the above algorithm solves the problem of minimizing the modular function f. The problem of minimizing a modular function on a distributive lattice, or equivalently, the problem of finding a minimum-weight ideal of a partially ordered set can also be solved by the technique presented in [12]. If $\nu: \mathcal{P} \to R$ is a monotone nondecreasing function, then the set of minimizers of f is trivially given by an interval $[A_-, A_0]_{\mathcal{D}}$ of $\mathcal{D}$, where $A_- = \{e \mid e \in E, \nu(e) < 0\}$ and $A_0 = \{e \mid e \in E, \nu(e) \leqq 0\}$. This is a very special case of the results of [2] and [9] on the so-called greedy algorithm.

Remark 5.5: Once the expression (5.16) is obtained, any nonnegative linear combination of $ER(A)$ which expresses $x - \nu$ can be obtained by repeatedly changing the values of the coefficients along cycles and paths with both end-vertices in E^+ or E^-, where these cycles and paths should be chosen such that they do not contain any arc in $\{\Delta^-(A) \mid x \in \partial f(A)\}$.

Acknowledgments: The author would like to express his thanks to Bill Cunningham for drawing the author's attention to the paper [12] and to Martin Grötschel for suggesting the examination of the relationship between subdifferentials of a submodular function and of its Lovász's extension. The author is also grateful to the referees for their useful comments.

References

[1] J. Edmonds, "Submodular functions, matroids, and certain polyhedra," Proceedings of the Calgary International Conference on Combinatorial Structures and Their Applications (Gordon and Breach, New York, 1970), pp. 69-87.

[2] U. Faigle, "The greedy algorithm for partially ordered sets," Discrete Mathematics 28 (1979) 153-159.

[3] L. R. Ford, Jr. and D. R. Fulkerson, Flows in Networks (Princeton University Press, Princeton, N.J., 1962).

[4] S. Fujishige, "The independent-flow problems and submodular functions" (in Japanese), Journal of the Faculty of Engineering, University of Tokyo A-16 (1978) 42-43.

[5] S. Fujishige, "Structures of polyhedra determined by submodular functions on crossing families," (to appear in Mathematical Programming).

[6] S. Fujishige, "Theory of submodular programs: a Fenchel-type min-max theorem and subgradients of submodular functions," (to appear in Mathematical Programming).

[7] S. Fujishige and N. Tomizawa, "A note on submodular functions on distributive lattices," (to appear in Journal of the Operations Research Society of Japan).

[8] A. J. Hoffman, "Some recent applications of the theory of linear inequalities to extremal combinatorial analysis," Proc. Symp. Appl. Math. 10 (1960) 113-127.

[9] B. Korte and L. Lovász, "Mathematical structures underlying greedy algorithms," Report No. 81189-OR, Institut fur Okonometrie und Operations Research, Universitat Bonn (July 1981).

[10] L. Lovász, "Submodular functions and convexity," presented at the XI. International Symposium on Mathematical Programming, Bonn, August 23-27, 1982.

[11] C. J. H. McDiarmid, "Rado's theorem for polymatroids," Mathematical Proceedings of the Cambridge Philosophical Society 78 (1975) 263-281.

[12] J. C. Picard, "Maximal closure of a graph and applications to combinatorial problems," Management Science 22 (1976) 1268-1272.

[13] R. T. Rockafellar, Convex Analysis (Princeton University Press, Princeton, N. J., 1970).

[14] J. Stoer and C. Witzgall, Convexity and Optimization in Finite Dimensions I (Springer, Berlin, 1970).

[15] N. Tomizawa, "Theory of hyperspace (I) — supermodular functions and generalization of concept of 'bases'"(in Japanese), Papers of The Technical Group on Circuits and Systems, Institute of Electronics and Communication Engineers of Japan, CAS80-72 (September 1980).

[16] N. Tomizawa, unpublished note (July 1982), included in "Theory of hyperspace (XVI) — on the structures of hedrons" (in Japanese), ibid., CAS82-174 (March 1983).

[17] N. Tomizawa and S. Fujishige, "Theory of hyperspace (VIII) — on the structures of hypermatroids of network type" (in Japanese), ibid., CAS81-62 (September 1981).