No. 135 (82-2)

The Theory of Semimodular Programs

by

Satoru FUJISHIGE

January 1982

THE THEORY OF SEMIMODULAR PROGRAMS

Satoru FUJISHIGE

Istitute of Socio-Economic Planning
University of Tsukuba
Sakura, Ibaraki 305, JAPAN

Abstract — We consider semimodular programs which are problems of minimizing submodular functions (or maximizing supermodular functions) with or without constraints. We define a convex (or concave) conjugate function of a submodular (or supermodular) function and show Fenchel's duality theorem for semimodular functions. We also define a subgradient of a submodular function and derive a necessary and sufficient condition for a feasible solution of a semimodular program to be optimal, which is a counterpart of the Karush-Kuhn-Tucker condition for convex programs. Moreover, solution algorithms for semimodular programs are proposed.

1. Introduction

Let $\mathcal{D}$ be a distributive lattice formed by subsets of a finite set E relative to set inclusion and let f be a set function from $\mathcal{D}$ to the set R of reals such that

$$f(A) + f(B) \geqq f(A \cup B) + f(A \cap B)$$

for any $A, B \in \mathcal{D}$. Then f is called a submodular (or lower semimodular) function on $\mathcal{D}$. If $-g$ is a submodular function, then g is called a supermodular (or upper semimodular) function. We call h a semimodular function if h is a submodular or supermodular function. If f is a submodular and, at the same time, supermodular function, then we call f a modular function.

We shall develop the theory of semimodular programs from the point of view of the duality in mathematical programs. Submodular (or supermodular) functions on distributive lattices share similar structures with convex (or concave) functions on convex sets. We shall define a convex (or concave) conjugate function of a submodular (or supermodular) function and show Fenchel's duality theorem for semimodular functions. We also define a subgradient of a submodular function and consider the problem of minimizing a submodular function with or without constraints. We derive a necessary and sufficient condition for a feasible solution to be optimal, which is a counterpart of the Karush-Kuhn-Tucker condition for convex programs.

2. Preliminaries

In this section we give propositions which are well known, or are immediate consequences of those well known, in polymatroid theory (see [2], [3], [7] and [10]).

Let E be a finite set, 2^E be the set of all the subsets of E and R be the set of reals. Throughout the present paper we assume, for the sake of simplicity, that for every distributive lattice $\mathcal{D} \subseteqq 2^E$ and every set function f defined on $\mathcal{D}$ we have $\emptyset \in \mathcal{D}$ and $f(\emptyset) = 0$.

For a distributive lattice $\mathcal{D} \subseteqq 2^E$ let $f:\mathcal{D} \to R$ be a submodular function. Then we call the pair $(\mathcal{D},f)$ a <u>submodular system</u>. Similarly, we define a supermodular system $(\mathcal{D},g)$ for a supermodular function $g:\mathcal{D} \to R$.

For distributive lattices $\mathcal{D}_1$, $\mathcal{D}_2 \subseteqq 2^E$, a submodular function $f:\mathcal{D}_1 \to R$ and a supermodular function $g:\mathcal{D}_2 \to R$, let us define polytopes

$$\mathbf{P}(f) = \{x \mid x \in R^E,\ \forall A \in \mathcal{D}_1 : x(A) \leqq f(A)\}, \tag{2.1}$$

$$\mathbf{P}(g) = \{x \mid x \in R^E,\ \forall A \in \mathcal{D}_2 : x(A) \geqq g(A)\} \tag{2.2}$$

and, if $E \in \mathcal{D}_1$ and $E \in \mathcal{D}_2$,

$$\mathbf{B}(f) = \{x \mid x \in \mathbf{P}(f), x(E)=f(E)\}, \tag{2.3}$$

$$\mathbf{B}(f) = \{x \mid x \in \mathbf{P}(g), x(E)=g(E)\}, \tag{2.4}$$

where R^E is the set of all $|E|$-vectors $x = (x(e): e \in E)$ with $x(e) \in R$ $(e \in R)$ and for any $A \subseteqq E$

$$x(A) = \sum_{e \in A} x(e). \tag{2.5}$$

Each vector $x \in R^E$ will also be regarded as a modular function $x:2^E \to R$ by (2.5). We call $\mathbf{P}(f)$ the <u>submodular polytope</u> associated with the submodular system $(\mathcal{D}_1, f)$ and $B(f)$ the <u>base polytope</u> associated with $(\mathcal{D}_1, f)$. Similarly, we call $P(g)$ the <u>supermodular polytope</u> associated with the supermodular system $(\mathcal{D}_2, g)$ and $B(g)$ the <u>base polytope</u> associated with $(\mathcal{D}_2, g)$. A vector $x \in B(f)$ (or $B(g)$) is called a <u>base</u> of $(\mathcal{D}_1, f)$ (or $(\mathcal{D}_2, g)$).

<u>Proposition 2.1</u>: For any submodular system $(\mathcal{D}, f)$ and $y \in P(f)$, if $E \in \mathcal{D}$, then there exists a base $x \in B(f)$ such that $y \leqq x$, i.e., $\forall e \in E$: $y(e) \leqq x(e)$.

<u>Proposition 2.2</u>: For any submodular system $(\mathcal{D}, f)$ and $A \in \mathcal{D}$ there exists a vector $x \in P(f)$ such that $x(A) = f(A)$.

For a distributive lattice $\mathcal{D} \subseteq 2^E$ with $E \in \mathcal{D}$ let us denote by $\overline{\mathcal{D}}$ the <u>dual</u> distributive lattice of $\mathcal{D}$ given by

$$\overline{\mathcal{D}} = \{E - A \mid A \in \mathcal{D}\}. \tag{2.6}$$

For a submodular function $f: \mathcal{D} \to R$ with $E \in \mathcal{D}$ define a supermodular function $g: \overline{\mathcal{D}} \to R$ by

$$g(E - A) = f(E) - f(A) \qquad (A \in \mathcal{D}). \tag{2.7}$$

We call g the <u>dual supermodular function</u> of f and denote it by $f^{\#}$. Similarly, for a supermodular function $g: \mathcal{D} \to R$ with $E \in \mathcal{D}$ we define the <u>dual submodular function</u> $g^{\#}$ of g. Note that $(f^{\#})^{\#} = f$ and $(g^{\#})^{\#} = g$.

Proposition 2.3: Suppose that $\mathcal{D} \subseteq 2^E$ is a distributive lattice with $\emptyset$, $E \in \mathcal{D}$ and $f:\mathcal{D} \to R$ is a submodular function. Then we have $\mathbf{B}(f) = \mathbf{B}(f^{\#})$.

Furthermore, for a submodular system $(\mathcal{D},f)$ let us define a function $r_f:R^E \to R$ by

$$r_f(x) = \max\{y(E) \mid y \leqq x, y \in \mathbf{P}(f)\} \tag{2.8}$$

for each $x \in R^E$. We call r_f the vector rank function of $(\mathcal{D},f)$.

Proposition 2.4: For a vector rank function r_f of a submodular system $(\mathcal{D},f)$ we have

$$r_f(x) = \min\{f(A) + x(E - A) \mid A \in \mathcal{D}\} \tag{2.9}$$

for each $x \in R^E$.

The following proposition generalizes Proposition 2.4.

Proposition 2.5: Let $f_1:\mathcal{D}_1 \to R$ and $f_2:\mathcal{D}_2 \to R$ be submodular functions and suppose that there exists a set $A \in \mathcal{D}_1$ such that $E - A \in \mathcal{D}_2$. Then we have

$$\min\{f_1(A) + f_2(E - A) \mid A \in \mathcal{D}_1, E - A \in \mathcal{D}_2\}$$
$$= \max\{x(E) \mid x \in \mathbf{P}(f_1) \cap \mathbf{P}(f_2)\}. \tag{2.10}$$

Corollary 2.6: For any submodular function $f:\mathcal{D} \to R$,

$$\min\{f(A) \mid A \in \mathcal{D}\} = \max\{x(E) \mid x \leqq 0, x \in \mathbf{P}(f)\}. \tag{2.11}$$

Every submodular (or supermodular) function yields a distributive lattice as follows.

Proposition 2.7: For any submodular (or supermodular) function $f: \mathcal{D} \to R$ (or $g: \mathcal{D} \to R$) the set of all the minimizers of f (or the maximizers of g) forms a distributive sublattice of $\mathcal{D}$.

3. Conjugate Functions of Semimodular Functions and Fenchel's Duality Theorem for Semimodular Functions

Let $f: \mathcal{D} \to R$ and $g: \mathcal{D} \to R$, respectively, be a submodular function and a supermodular function. Let us define a function $f^*: R^E \to R$ by

$$f^*(x) = \max\{x(A) - f(A) \mid A \in \mathcal{D}\} \tag{3.1}$$

and, similarly, a function $g^*: R^E \to R$ by

$$g^*(x) = \min\{x(A) - g(A) \mid A \in \mathcal{D}\}. \tag{3.2}$$

It follows from (3.1) and (3.2) that f^* (or g^*) is a convex (or concave) function on R^E. We call f^* a convex conjugate function of f and g^* a concave conjugate function of g.

Theorem 3.1: For a submodular function $f: \mathcal{D} \to R$ and a supermodular function $g: \mathcal{D} \to R$ we have

$$f(A) = \max\{x(A) - f^*(x) \mid x \in R^E\}, \tag{3.3}$$

$$g(A) = \min\{x(A) - g^*(x) \mid x \in R^E\} \quad (3.4)$$

for any $A \in \mathcal{D}$.

(Proof) From (3.1) we have

$$f(A) \geqq x(A) - f^*(x) \quad (3.5)$$

for any $A \in \mathcal{D}$ and $x \in R^E$. Therefore, we shall show that for any $A \in \mathcal{D}$ there exists a vector $x \in R^E$ such that (3.5) holds with equality, which completes the proof of (3.3).

It follows from Proposition 2.2 that for any $A \in \mathcal{D}$ there exists a vector $\hat{x} \in R^E$ such that

$$f(B) \geqq \hat{x}(B) \quad (3.6)$$

for any $B \in \mathcal{D}$, where (3.6) holds with equality when $B = A$. Therefore, by the definition of f^* we have

$$f^*(x) = \hat{x}(A) - f(A) \ (= 0), \quad (3.7)$$

which implies (3.3).

The relation (3.4) can be shown in a similar way. Q.E.D.

We see from Theorem 3.1 that the correspondence between a submodular (or supermodular) function f (or g) and its convex (or concave) conjugate function f^* (or g^*) is one to one. Let us define

$$(f^*)^* = f, \qquad (g^*)^* = g. \quad (3.8)$$

For submodular functions $f_i : \mathcal{D}_i \rightarrow R$ $(i=1,2)$, define a <u>convolution</u> $f_1^* \circ f_2^*$ of f_1^* and f_2^* by

$$(f_1^* \circ f_2^*)(x) = \min\{f_1^*(x_1) + f_2^*(x_2) \mid x_1 + x_2 = x\}. \quad (3.9)$$

Note that the minimum in (3.9) exists for each $x \in R^E$.

Theorem 3.2: For submodular functions $f_i : \mathcal{D}_i \to R$ $(i=1,2)$ we have

$$(f_1^* \circ f_2^*)^* = f_1 + f_2, \tag{3.10}$$

where the domain of $f_1 + f_2$ is $\mathcal{D}_1 \cap \mathcal{D}_2$.

(Proof) By the definition (3.9) of convolution we have for any $A \in \mathcal{D}_1 \cap \mathcal{D}_2$

$$\begin{aligned}
&(f_1^* \circ f_2^*)^*(A) \\
&= \max\{x(A) - (f_1^* \circ f_2^*)(x) \mid x \in R^E\} \\
&= \max\{x(A) - \min\{f_1^*(x_1) + f_2(x_2) \mid x_1 + x_2 = x\} \mid x \in R^E\} \\
&= \max\{x_1(A) - f_1^*(x_1) + x_2(A) - f_2^*(x_2) \mid x_1, x_2 \in R^E\} \\
&= f_1(A) + f_2(A),
\end{aligned}$$

where the last equality follows from Theorem 3.1. Q.E.D.

It should be noted that Theorem 3.1 is closely related to Proposition 2.4.

Theorem 3.3: Let $f : \mathcal{D} \to R$ be a submodular function. Then for any $x \in R^E$ $f^*(x)$ is given by

$$f^*(x) = x(E) - r_f(x), \tag{3.11}$$

where r_f is the vector rank function associated with $(\mathcal{D}, f)$.

(Proof) The theorem immediately follows from Proposition 2.4 and the definition (3.1) of f^*. Q.E.D.

Now, we show Fenchel's duality theorem for semimodular functions. (For Fenchel's duality theorem for convex and concave functions, see [8] and [9].)

<u>Theorem 3.4</u>: For a submodular function $f:\mathcal{D} \to R$ and a supermodular function $g:\mathcal{D} \to R$ we have

$$\min\{f(A) - g(A) \mid A \in \mathcal{D}\} = \max\{g^*(x) - f^*(x) \mid x \in R^E\}. \quad (3.12)$$

(Proof) Without loss of generality we can suppose $E \in \mathcal{D}$. Then the min = max relation (3.12) is equivalent to

$$\min\{f(A) + g^{\#}(E - A) \mid A \in \mathcal{D}\}$$
$$= \max\{g^{\#}(x) + g(E) - f^*(x) \mid x \in R^E\}, \quad (3.13)$$

where $g^{\#}$ is the dual submodular function of g. Moreover, from Theorem 3.3 and Proposition 2.4 we have

$$f^*(x) = x(E) - r_f(x), \quad (3.14)$$

$$g^*(x) + g(E) = \min\{x(A) + g^{\#}(E - A) \mid A \in \mathcal{D}\}$$
$$= r_{g^{\#}}(x). \quad (3.15)$$

It follows from (3.14) and (3.15) that (3.13) can be rewritten as

$$\min\{f(A) + g^{\#}(E - A) \mid A \in \mathcal{D}\}$$
$$= \max\{r_f(x) + r_{g^{\#}}(x) - x(E) \mid x \in R^E\}. \quad (3.16)$$

For any $x \in R^E$ let y and z be vectors in R^E satisfying

$$y \in \mathbf{P}(f), \quad y \leqq x, \quad y(E) = r_f(x), \quad (3.17)$$

$$z \in \mathbf{P}(g^{\#}), \quad z \leqq x, \quad z(E) = r_{g^{\#}}(x). \quad (3.18)$$

Define

$$\bar{x} = y \wedge z \equiv (\min(y(e), z(e)) : e \in E). \quad (3.19)$$

Then we have

$$\bar{x} \in \mathbf{P}(f) \cap \mathbf{P}(g^{\#}). \quad (3.20)$$

Because of (3.19) and (3.20),

$$r_f(x) + r_{g^{\#}}(x) - x(E)$$
$$= y(E) + z(E) - x(E)$$

$$= \overline{x}(E) + \overline{x}(E) - \overline{x}(E)$$

$$= r_f(\overline{x}) + r_{g^\#}(\overline{x}) - \overline{x}(E) \quad (= \overline{x}(E)). \tag{3.21}$$

Therefore, we see from (3.20) and (3.21) that (3.16) is equivalent to

$$\min\{f(A) + g^\#(E - A) \mid A \in \mathcal{D}\}$$

$$= \max\{x(E) \mid x \in \mathbf{P}(f) \cap \mathbf{P}(g^\#)\}, \tag{3.22}$$

which is valid due to Proposition 2.5. Q.E.D.

The proof of Theorem 3.4 reveals that Theorem 3.4 is equivalent to Proposition 2.5. However, the min = max relation in the form of Theorem 3.4 motivates further investigation of semimodular functions from the point of view of the duality theory of convex programs [8], [9].

4. Subgradients of Submodular Functions

Let $f:\mathcal{D} \to R$ be a submodular function. For a vector $x \in R^E$ and a set $A \in \mathcal{D}$, if

$$f(B) - f(A) \geq x(B) - x(A) \tag{4.1}$$

holds for any $B \in \mathcal{D}$, then we call x a subgradient of f at A. We denote by $\partial f(A)$ the set of all subgradients of f at $A \in \mathcal{D}$ and call $\partial f(A)$ the subdifferential of f at A. It should be noted that R^E is divided into $|\mathcal{D}|$ parts $\partial f(A)$ $(A \in \mathcal{D})$.

Lemma 4.1: For a submodular function $f:\mathcal{D} \to R$, a vector $x \in R^E$ and a set $A \in \mathcal{D}$, the following (i), (ii) and (iii) are equivalent to one another:

$$\text{(i)}\quad x \in \partial f(A); \tag{4.2}$$

$$\text{(ii)}\quad \min\{f(B) + x(E - B) \mid B \in \mathcal{D}\} = f(A) + x(E - A); \tag{4.3}$$

$$\text{(iii)}\quad f(A) + f^*(x) = x(A). \tag{4.4}$$

(Proof) The lemma immediately follows from the definition of $\partial f(A)$.

Q.E.D.

<u>Lemma 4.2</u>: For a submodular function $f:\mathcal{D} \to R$ we have

$$\text{(a)}\quad \partial f(\emptyset) = \mathbf{P}(f); \tag{4.5}$$

$$\text{(b)}\quad \text{if } E \in \mathcal{D},\ \partial f(E) = \mathbf{P}(f^{\#}); \tag{4.6}$$

$$\text{(c)}\quad \text{if } E \in \mathcal{D},\ \text{for each } A \in \mathcal{D}\quad \partial f(A) \cap \mathbf{B}(f) \neq \emptyset, \tag{4.7}$$

where $f^{\#}$ is the dual supermodular function of f.

(Proof) The lemma follows from Proposition 2.2 and the definitions of $f^{\#}$ and $\partial f(A)$ $(A \in \mathcal{D})$. Q.E.D.

For a convex conjugate function f^* of a submodular function $f:\mathcal{D} \to R$ and a vector $x \in R^E$ let us define a set $\partial f^*(x)$ of subsets of E as follows:

$$A \in \partial f^*(x) \tag{4.8}$$

if and only if $A \subseteqq E$ and

$$f^*(y) - f^*(x) \geqq y(A) - x(A) \tag{4.9}$$

for any $y \in R^E$. We call $\partial f^*(x)$ the <u>binary subdifferential</u> of f^*.

Theorem 4.3: For a submodular function $f:\mathcal{D} \to R$ we have

$$f^*(x) \subseteq \mathcal{D} \tag{4.10}$$

for each $x \in R^E$. Moreover, for a vector $x \in R^E$ and a set $A \in \mathcal{D}$ the following (i) and (ii) are equivalent to each other:

$$\text{(i)} \quad x \in \partial f(A); \tag{4.11}$$

$$\text{(ii)} \quad A \in \partial f^*(x). \tag{4.12}$$

(Proof) The relation (4.10) easily follows from the definition of $\partial f^*(x)$. Furthermore, from Theorem 3.1 and (4.9) we see that (4.12) is equivalent to

$$\begin{aligned} x(A) - f^*(x) &= \max\{y(A) - f^*(y) \mid y \in R^E\} \\ &= f(A). \end{aligned} \tag{4.13}$$

It follows from Lemma 4.1 that (4.13) is equivalent to (4.11). Q.E.D.

Lemma 4.4: For a submodular function $f:\mathcal{D} \to R$ the binary subdifferential $\partial f^*(x)$ forms a distributive sublattice of $\mathcal{D}$ for each $x \in R^E$.

(Proof) The lemma follows from Proposition 2.7 and the fact that " $A \in \partial f^*(x)$ " is equivalent to " A is a maximizer of a supermodular function $x - f: \mathcal{D} \to R$." Q.E.D.

Let us denote by $\hat{\partial} f^*(x)$ the subdifferential of the convex conjugate function f^* at $x \in R^E$ in an ordinary sense (cf. [8]). Let us call $\hat{\partial} f^*(x)$ the <u>real subdifferential</u> of f^* at x.

Theorem 4.5: For a submodular function $f:\mathcal{D} \to R$, the real subdifferential

$\hat{\partial} f^*(x)$ of f^* at $x \in R^E$ is a convex polytope and the set of all the extreme points of $\hat{\partial} f^*(x)$ is given by $\{\chi_A \mid A \in \partial f^*(x)\}$, where $\chi_A \in R^E$ is the characteristic vector of a set A.

(Proof) Since f^* is the point-wise maximum of the linear functions $x(A) - f(A)$ of x $(A \in \mathcal{D})$, f^* is a polyhedral convex function and the theorem easily follows. We omit the detail. Q.E.D.

Theorem 4.6: Suppose that $f_i : \mathcal{D}_i \to R$ is a submodular function on a distributive lattice $\mathcal{D}_i$ for each $i = 1, 2, \ldots, m$. Let $f : \mathcal{D} \equiv \bigcap_{i=1}^{m} \mathcal{D}_i \to R$ be a submodular function defined by

$$f(A) = f_1(A) + f_2(A) + \cdots + f_m(A) \tag{4.14}$$

for $A \in \mathcal{D}$. Then for any $A \in \mathcal{D}$ we have

$$\partial f(A) = \partial f_1(A) + \partial f_2(A) + \cdots + \partial f_m(A). \tag{4.15}$$

(Proof) We see from Lemma 4.1 that

$$x \in \partial f(A) \tag{4.16}$$

if and only if

$$f(A) + f^*(x) = x(A). \tag{4.17}$$

From (4.17) and Theorem 3.2 we have

$$f(A) + \min\{f_1{}^*(x_1) + f_2{}^*(x_2) + \cdots + f_m{}^*(x_m) \mid x_1 + x_2 + \cdots + x_m = x\} = x(A). \tag{4.18}$$

Let $\hat{x}_i$ $(i=1,2,\ldots,m)$ be vectors by which the minimum in (4.18) is attained. Then (4.18) is rewritten as

$$f_1(A) + f_1{}^*(\hat{x}_1) + f_2(A) + f_2{}^*(\hat{x}_2) + \cdots + f_m(A) + f_m{}^*(\hat{x}_m) = \hat{x}_1(A) + \hat{x}_2(A) + \cdots + \hat{x}_m(A). \tag{4.19}$$

Since for each $i = 1, 2, \ldots, m$ we have

$$f_i(A) + f_i^*(\hat{x}_i) \geqq \hat{x}_i(A), \qquad (4.20)$$

(4.19) implies that (4.20) holds with equality for each $i = 1, 2, \ldots, m$. Consequently, (4.16) is equivalent to the following (4.21) and (4.22):

$$\hat{x}_i \in \partial f_i(A) \quad (i = 1, 2, \ldots, m), \qquad (4.21)$$

$$x = \hat{x}_1 + \hat{x}_2 + \cdots + \hat{x}_m. \qquad (4.22)$$

Q.E.D.

5. Submosular Programs

In this section we consider the problem of minimizing submodular functions with or without constraints.

5.1 Unconstrained Minimization of Submodular Functions

The definition of a subdifferential of a submodular function directly yields the following fundamental lemma.

Lemma 5.1: Let $f:\mathcal{D} \to R$ be a submodular function. The minimum value of f is attained by $A \in \mathcal{D}$ if and only if

$$0 \in \partial f(A), \qquad (5.1)$$

where 0 is the zero vector in R^E.

Consequently, the problem of minimizing a submodular function $f:\mathcal{D} \to R$ is equivalent to that of finding a set $A \in \mathcal{D}$ such that

(5.1) holds.

In this section we assume that $\emptyset$, $E \in \mathcal{D}$ and that for each $e \in E$ there exist $A, B \in \mathcal{D}$ such that $A \subseteqq B$ and $B - A = \{e\}$. This assumption is only for the sake of simplisity. Then there exists a partially ordered set $\mathcal{P} = (E, \preccurlyeq)$ and the distributive lattice $\mathcal{D}$ can be expressed as

$$\mathcal{D} = \{D^- \mid (D^-, D^+) \text{ is a monotone dissection of } \mathcal{P}\}, \qquad (5.2)$$

where (D^-, D^+) is a monotone dissection of $\mathcal{P} = (E, \preccurlyeq)$ if and only if $D^- \cup D^+ = E$, $D^- \cap D^+ = \emptyset$ and for any $e_1, e_2 \in E$ " $e_1 \preccurlyeq e_2$ and $e_2 \in D^-$ " implies " $e_1 \in D^-$ ". The one-to-one correspondence between the set of distributive lattices $\mathcal{D} \subseteqq 2^E$ and the set of partially ordered sets $\mathcal{P} = (E, \preccurlyeq)$ on E is pointed out in [1].

Let us define $C(e) \in \mathcal{D}$ $(e \in E)$ by

$$C(e) = \cup\{A \mid e \notin A \in \mathcal{D}\} \qquad (5.3)$$

and a vector $\acute{f} \in R^E$ by

$$\acute{f}(e) = f(C(e) \cup \{e\}) - f(C(e)). \qquad (5.4)$$

Here, it should be noted that $C(e) \cup \{e\} \in \mathcal{D}$ $(e \in E)$. When $\mathcal{D} = 2^E$, $C(e) = E - \{e\}$ and the vector $\acute{f}$ is called the <u>greater lower bound</u> of $B(f)$ in [10].

<u>Lemma 5.2</u>: Under the above assumptions the function $f - \acute{f}: \mathcal{D} \to R$ is monotone nondecreasing.

(Proof) It is sufficient to show that, for $A, B \in \mathcal{D}$ such that $\{e\} = B - A$ and $B = A \cup \{e\}$, we have

$$f(B) - \acute{f}(B) \geqq f(A) - \acute{f}(A). \qquad (5.5)$$

It follows from (5.4) and the submodularity of $f - \acute{f}$ that

$$f(B) - \acute{f}(B) - (f(A) - \acute{f}(A))$$
$$= f(B) + f(C(e)) - f(C(e) \cup \{e\}) - f(A)$$
$$\geqq 0, \qquad (5.6)$$

where note that $C(e) \cup \{e\} = B \cup C(e)$ and $A = B \cap C(e)$. Q.E.D.

Because of Lemma 5.2 we consider the problem formulated as follows:

"Given a monotone nondecreasing submodular function $f:\mathcal{D} \to R$ and a positive vector $w \in R^E$, find a set $A \in \mathcal{D}$ such that $w \in \partial f(A)$." (5.7)

The original problem can be transformed into the problem (5.7) by putting

$$w \leftarrow - \acute{f} + u, \qquad (5.8)$$
$$f \leftarrow f + w, \qquad (5.9)$$

where $u \in R^E$ is any nonnegative vector such that w given by (5.8) is a positive vector.

<u>Theorem 5.3</u>: For f and w in (5.7) there exists a unique base $\hat{x} \in \mathrm{B}(f)$ such that, for each $\alpha \geqq 0$, a vector $x \in R^E$ given by

$$x = \hat{x} \wedge \alpha w \qquad (5.10)$$

is a maximal vector with the property that $x \in \mathrm{P}(f)$ and $x \leqq \alpha w$, or

$$\min\{f(A) + \alpha w(E - A) \mid A \in \mathcal{D}\} = (\hat{x} \wedge \alpha w)(E). \qquad (5.11)$$

Moreover, let the distinct values of $\hat{x}(e)/w(e)$ $(e \in E)$ be given by

$$c_1 < c_2 < \cdots < c_p \tag{5.12}$$

and define

$$T_i = \{e \mid e \in E, \hat{x}(e)/w(e) \leqq c_i\} \quad (i=1,2,\ldots,p). \tag{5.13}$$

Also let i_0 be an integer such that $1 \leqq i_0 \leqq p$ and

$$c_{i_0} \leqq 1 < c_{i_0+1}, \tag{5.14}$$

where we put $c_{p+1} = +\infty$. Then we have

$$w \in \partial f(T_{i_0}). \tag{5.15}$$

The base $\hat{x}$ in Theorem 5.3 for the case when $\mathcal{D} = 2^E$ is called a <u>lexicographically</u> <u>optimal</u> <u>base</u> of a polymatroid (E,f) with respect to a weight vector w [4]. Theorem 5.3 can be shown by direct adaptation of the argument in [4] and the proof is omitted.

For a set $A \in \mathcal{D}$ and a vector $x \in R^E$ we call A a <u>hyperplane</u> <u>separating</u> $P(f)$ $(= \partial f(\emptyset))$ and x if $f(A) < x(A)$.

The value c_1 in (5.12) is characterized by the following (a) and (b):

(a) there exists no hyperplane separating $P(f)$ and $c_1 w$, i.e., $c_1 w \in P(f)$; (5.16)

(b) for any $\alpha > c_1$ there exists a hyperplane separating $P(f)$ and αw, i.e., $\alpha w \notin P(f)$. (5.17)

Moreover, T_1 in (5.13) is given by

$$T_1 = \{e \mid e \in E,\ \forall\, \varepsilon > 0\colon c_1 w + \varepsilon \chi_{\{e\}} \notin P(f)\}, \tag{5.18}$$

where $\chi_{\{e\}} \in R^E$ is the characteristic vector of $\{e\}$.

If f is integer-valued, then c_1 and T_1 can be obtained by a binary search based on (5.16) - (5.18) using an oracle for discerning the existence of a separating hyperplane. Here, it should be noted that, for a distributive lattice $\mathcal{D}_1$ given by

$$\mathcal{D}_1 = \{A \mid A \in \mathcal{D}, f(A) = c_1 w(A)\}, \tag{5.19}$$

T_1 is the maximum element of $\mathcal{D}_1$ and

$$\forall A \in \mathcal{D}_1 : c_1 = f(A)/w(A). \tag{5.20}$$

Therefore, for a sufficiently small $\varepsilon > 0$ every hyperplane $A \in \mathcal{D}$ separating $\mathbf{P}(f)$ and $(c_1 + \varepsilon)w$ belongs to $\mathcal{D}_1$ and gives us the value c_1 by (5.20). It may also be noted that

$$c_p < \max\{f(D(e))/w(D(e)) \mid e \in E\}, \tag{5.21}$$

where

$$D(e) = \cap\{A \mid e \in A \in \mathcal{D}\} \quad (e \in E). \tag{5.22}$$

Once c_1 and T_1 are obtained, contract T_1, i.e, consider $f/T_1 : \mathcal{D}/T_1 \to R$ defined by

$$\mathcal{D}/T_1 = \{A \mid A \subseteq E - T_1, A \cup T_1 \in \mathcal{D}\}, \tag{5.23}$$

$$(f/T_1)(A) = f(A \cup T_1) - f(T_1) \quad (A \in \mathcal{D}/T_1), \tag{5.24}$$

instead of f and repeat the above process until we find i_0 satisfying (5.14).

We have proposed an algorithm for minimizing a submodular function $f : \mathcal{D} \to R$ using a separating-hyperplane oracle. Another algorithm using a separating-hyperplane oracle was proposed by M. Grötschel, L. Lovász and A. Schrijver [6], where a binary search is adopted in the range of f. Our algorithm is a polynomial-time one in the sense of [6] as well.

5.2 Constrained Minimization of Submodular Functions

Let $\mathcal{D}_0 \subseteq 2^E$ be a distributive lattice. We say that a vector $x \in R^E$ is normal to $\mathcal{D}_0$ at $A \in \mathcal{D}_0$ if for each $B \in \mathcal{D}_0$ we have

$$x(B) - x(A) \leq 0. \tag{5.25}$$

Theorem 5.4: Let $\mathcal{D}$ and $\mathcal{D}_0$ be distributive lattices with $\mathcal{D}_0 \subseteq \mathcal{D} \subseteq 2^E$. Also let $f: \mathcal{D} \to R$ be a submodular function. Then we have

$$f(A) = \min\{f(B) \mid B \in \mathcal{D}_0\} \tag{5.26}$$

for $A \in \mathcal{D}_0$ if and only if there exists a subgradient $x \in \partial f(A)$ such that $-x$ is normal to $\mathcal{D}_0$ at A.

(Proof) The "if" part: From the assumption we have

$$f(B) - f(A) \geq x(B) - x(A) \geq 0 \tag{5.27}$$

for any $B \in \mathcal{D}_0$.

The "only if" part: Define a (sub)modular function $f_0: \mathcal{D}_0 \to R$ by

$$f_0(B) = 0 \quad (B \in \mathcal{D}_0). \tag{5.28}$$

Then A is a minimizer of $f_1 \equiv f + f_0$: $\mathcal{D}_0 \to R$. It follows from Theorem 4.6 and Lemma 5.1 that

$$0 \in \partial f_1(A) = \partial f(A) + \partial f_0(A). \tag{5.29}$$

Therefore, there exists a vector $x \in \partial f(A)$ such that

$$- x \in \partial f_0(A). \tag{5.30}$$

By the definition of f_0, (5.30) implies that $-x$ is normal to $\mathcal{D}_0$ at A. Q.E.D.

Suppose that for each $i = 0, 1, \ldots, m$ $f_i : \mathcal{D} \to R$ is a submodular function and the minimum value of f_i is equal to α_i.

Let us consider a constrained minimization problem as follows.

(P): Minimize $f_0(A)$

subject to $f_i'(A) = 0$ $(i=1,2,\ldots,m)$. (5.31)

where $f_i'(A) = f_i(A) - \alpha_i$ $(i=1,2,\ldots,m)$. Note that $\mathcal{D}_0 \subseteqq 2^E$ given by

$$\mathcal{D}_0 = \{A \mid \forall\, i=1,2,\ldots,m:\ f_i'(A)=0\} \tag{5.32}$$

is a distributive lattice.

Define a function $L: R_+^m \times \mathcal{D} \to R$ by

$$L(\lambda,A) = f_0(A) + \lambda_1 f_1'(A) + \cdots + \lambda_m f_m'(A), \tag{5.33}$$

where R_+ is the set of nonnegative reals, $\lambda = (\lambda_1,\ldots,\lambda_m) \in R_+^m$ and $A \in \mathcal{D}$. We call L a <u>Lagrangian function</u> associated with the above problem (P). We call $\hat{\lambda} \in R_+^m$ an <u>optimal Lagrange multiplier</u> if

$$\min\{L(\hat{\lambda},A) \mid A \in \mathcal{D}\} \tag{5.34}$$

is equal to the optimal value of the objective function of problem (P).

<u>Theorem 5.5</u>: For problem (P) given by (3.31),

(a-1) $\hat{\lambda} = (\hat{\lambda}_1,\ldots,\hat{\lambda}_m)$ is an optimal Lagrange multiplier

and

(a-2) $\hat{A}$ is an optimal solution of (P)

if and only if

(b-1) $\hat{\lambda} = (\hat{\lambda}_1,\ldots,\hat{\lambda}_m) \in R_+^m$,

(b-2) $\hat{A}$ is a feasible solution of (P)

and

(b-3) $0 \in \partial f_0(\hat{A}) + \hat{\lambda}_1 \partial f_1(\hat{A}) + \cdots + \hat{\lambda}_m \partial f_m(\hat{A})$.

(Proof) <u>The "if" part</u>: From (b-1), (b-2) and (b-3) we have

$$\min\{L(\hat{\lambda},A) \mid A \in \mathcal{D}\} = L(\hat{\lambda},\hat{A}) = f_0(\hat{A}), \tag{5.35}$$

while, generally, we have

$$\min\{L(\hat{\lambda},A) \mid A \in \mathcal{D}\} \leqq f_0(B) \tag{5.36}$$

for any feasible solution B. (a-1) and (a-2) thus follows.

<u>The "only if" part</u>: From (a-1) and (a-2) we have $A \in \mathcal{D}_0$ and

$$\min\{L(\hat{\lambda},A) \mid A \in \mathcal{D}\} = f_0(\hat{A}) = L(\hat{\lambda},\hat{A}). \tag{5.37}$$

Therefore, we have (b-3). (b-1) and (b-2) trivially follows from (a-1) and (a-2). Q.E.D.

Let us define a function $p:R^m_+ \to R \cup \{+\infty\}$ by

$$p(u) = \min\{f_0(A) \mid \forall\ i=1,2,\ldots,m:\ f_i'(A) \leqq u_i\}, \tag{5.38}$$

where $u = (u_1,u_2,\ldots,u_m) \in R^m_+$. If for $u \in R^m_+$ there exists no $A \in \mathcal{D}$ such that $f_i'(A) \leqq u_i$ for all $i = 1, 2, \ldots, m$, then we define $p(u) = +\infty$. We call p a <u>perturbation function</u> associated with problem (P).

<u>Theorem 5.6</u>: Suppose that problem (P) has a feasible solution. For the perturbation function p associated with problem (P) we have for each $\lambda \in R^m_+$

$$\min\{p(u) + \lambda_1 u_1 + \cdots + \lambda_m u_m \mid u \in R^m_+\}$$
$$= \min\{L(\lambda,A) \mid A \in \mathcal{D}\}. \tag{5.39}$$

(Proof) Suppose that $\hat{A} \in \mathcal{D}$ is a minimizer of $L(\lambda,A)$ in $A \in \mathcal{D}$. It follows from the definition of $p(u)$ that

$$L(\lambda,\hat{A}) = p(\bar{u}) + \lambda_1\bar{u}_1 + \cdots + \lambda_m\bar{u}_m, \tag{5.40}$$

where $\bar{u}_i = f_i(\hat{A})$ $(i=1,\ldots,m)$. Consequently,

$$\min\{p(u) + \lambda_1 u_1 + \cdots + \lambda_m u_m \mid u \in R^m_+\}$$
$$\leq \min\{L(\lambda,A) \mid A \in \mathcal{D}\}. \tag{5.41}$$

On the other hand, suppose that $\hat{u} \in R^m_+$ is a minimizer of $p(u) + \lambda_1 u_1 + \cdots + \lambda_m u_m$ in $u \in R^m_+$. There exists an $A_0 \in \mathcal{D}$ such that

$$p(\hat{u}) = f_0(A_0), \tag{5.42}$$

$$f_i'(A_0) \leqq \hat{u}_i \qquad (i=1,\ldots,m). \tag{5.43}$$

Since $\lambda \in R^m_+$, we have from (5.42) and (5.43)

$$p(\hat{u}) + \lambda_1 \hat{u}_1 + \cdots + \lambda_m \hat{u}_m \geqq L(\lambda, A_0). \tag{5.44}$$

Therefore, we have

$$\min\{p(u) + \lambda_1 u_1 + \cdots + \lambda_m u_m \mid u \in R^m_+\}$$
$$\geqq \min\{L(\lambda,A) \mid A \in \mathcal{D}\}. \tag{5.45}$$

The theorem follows from (5.41) and (5.45). Q.E.D.

In the proof of Theorem 5.6 we have already shown the following.

Theorem 5.7: For $\lambda \in R^m_+$ and the perturbation function p associated with problem (P), we have

(a) if $\hat{A} \in \mathcal{D}$ is a minimizer of $L(\lambda,A)$ in $A \in \mathcal{D}$, then $\hat{u} = (\hat{u}_1,\ldots,\hat{u}_m)$ given by

$$\hat{u}_i = f_i(\hat{A}) \qquad (i=1,\ldots,m) \tag{5.46}$$

is a minimizer of $p(u) + \lambda_1 u_1 + \cdots + \lambda_m u_m$ in $u \in R^m_+$;

(b) if $\hat{u} = (\hat{u}_1,\ldots,\hat{u}_m)$ is a minimizer of $p(u) + \lambda_1 u_1 + \cdots + \lambda_m u_m$ in $u \in R^m_+$ and $\hat{A} \in \mathcal{D}$ satisfies

$$p(\hat{u}) = f_0(\hat{A}), \tag{5.47}$$

$$f_i'(\hat{A}) \leqq \hat{u}_i \quad (i=1,\ldots,m), \tag{5.48}$$

then $\hat{A}$ is a minimizer of $L(\lambda,A)$ in $A \in \mathcal{D}$ and for each $i = 1, \ldots, m$ (5.48) holds with equality if $\lambda_i > 0$.

Theorem 5.8: A vector $\hat{\lambda} \in R_+^m$ is an optimal Lagrange multuplier of problem (P) if and only if

$$\min\{p(u) + \hat{\lambda}_1 u_1 + \cdots + \hat{\lambda}_m u_m \mid u \in R_+^m\} = p(0). \tag{5.49}$$

(Proof) (5.49) means that the zero vector $0 \in R_+^m$ is a minimizer of $p(u) + \hat{\lambda}_1 u_1 + \cdots + \hat{\lambda}_m u_m$ in $u \in R_+^m$. Therefore, the theorem easily follows from Theorems 5.5 and 5.7. Q.E.D.

We see from Theorem 5.8 that, if $\hat{\lambda} \in R_+^m$ is an optimal Lagrange multiplier, then any $\lambda \in R_+^m$ with $\hat{\lambda} \leqq \lambda$ is also an optimal one.

An algorithm for solving problem (P) is given as follows. Recall that we have assumed that for each $i = 0, 1, \ldots, m$ the minimum value of f_i is equal to α_i.

We consider the case when $m = 1$.

An Algorithm for Solving Problem (P) (with $m = 1$)

Step 0°: Let a_0 be an upper bound of f_0 such that $a_0 \neq f(A)$ for all $A \in \mathcal{D}$. Also let a_1 be an upper bound of f_1. Put $\lambda \leftarrow (a_0 - \alpha_0)/(a_1 - \alpha_1)$.

Step 1°: Put $\hat{A} \leftarrow$ a minimizer of $L(\lambda,A) = f_0(A) + \lambda f_1'(A)$ in $A \in \mathcal{D}$.

Step 2°: If $f_1'(\hat{A}) = 0$, then stop ($\hat{A}$ is an optimal solution). Otherwise, put $\lambda \leftarrow (a_0 - f_0(\hat{A}))/f_1'(\hat{A})$ and go back to Step 1°.

The validity of the above algorithm follows from Theorem 5.5, 5.7 and 5.8. The case when $m > 1$ can also be treated by the algorithm by setting $f_1 \leftarrow f_1 + f_2 + \cdots + f_m$. If $\min\{f_1(A) + f_2(A) + \cdots + f_m(A) \mid A \in \mathcal{D}\} \neq \alpha_1 + \alpha_2 + \cdots + \alpha_m$, then there exists no feasible solution.

Upper bounds of submodular functions required in Step 0° are obtained based on the following lemma.

Lemma 5.9: For a submodular function $f:\mathcal{D} \to R$ let $\grave{f} \in R^E$ be a vector given by

$$\grave{f}(e) = f(D(e)) - f(D(e) - \{e\}) \qquad (e \in E), \tag{5.50}$$

where $D(e)$ is defined by (5.22). Then for any $A \in \mathcal{D}$ we have

$$f(A) \leqq \grave{f}(A). \tag{5.51}$$

(Proof) The lemma is in a form dual to Lemma 5.2 and we omit the proof (see [5]). Q.E.D.

An upper bound a of $f:\mathcal{D} \to R$ is given by

$$a = \sum\{\check{f}(e) \mid e \in E, \check{f}(e) > 0\}. \tag{5.52}$$

6. Concluding Remarks

We have developed the theory of semimodular programs and revealed the similarity or analogy between submodular functions on distributive lattices and convex functions on convex sets. The theory in the present paper can further be developed and elaborated as the theory of convex programs has been done so far. It may also be interesting to examine the existing algorithms for convex programs and investigate the possibility of adapting them to semimodular programs.

Acknowledgments

The author would like to express his sincere thanks to Professor Masao Iri of the University of Tokyo and Dr. Nobuaki Tomizawa of the Tokyo Institute of Technology for useful conversations and stimulating discussions.

References

[1] G. Birkhoff, Lattice Theory (American Mathematical Society Colloquium Publications 25, Third edition, Providence, RI, 1967).

[2] J. Edmonds, "Submodular functions, matroids, and certain polyhedra," In Proceedings of the International Conference on Combinatorial Structures and Their Applications (Gordon and Breach, New York, 1970) pp. 67-87.

[3] S. Fujishige, "Algorithms for solving the independent-flow problems," Journal of the Operations Research Society of Japan 21 (1978) 189-204.

[4] S. Fujishige, "Lexicographically optimal base of a polymatroid with respect to a weight vector," Mathematics of Operations Research 5 (1980) 186-196.

[5] S. Fujishige and N. Tomizawa, "A note on submodular functions on distributive lattices" (submitted).

[6] M. Grötschel, L. Lovász and A. Schrijver, "The ellipsoid method and its consequences in combinatorial optimization," Combinatorica 1 (1981) 169-197.

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

[8] R. T. Rockafellar, Convex Analysis (Princeton University Press, Princeton, NJ, 1970).

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

[10] N. Tomizawa, "Theory of hyperspace (I)-(XII)" (in Japanese), In Proceedings of the Research Meeting on Circuits and Systems (The Institute of Electronics and Communication Engineers of Japan, 1980-1981).

INSTITUTE OF SOCIO-ECONOMIC PLANNING

Discussion Paper Series (1981.4-   )

No. 113 Sho-Ichiro Kusumoto,
"On the Equilibrium Concepts in a General Equilibrium Theory with Public Goods and Taxes-Pareto Optimallity and Existence," (April, 1981).

No. 114 Ryosuke Hotaka,
"A Meta-Database for a Database Design Method," (May, 1981).

No. 115 Hidehiko Tanimura,
"A Minimum-Distance Location Model Central Facilities with Entropy-Maximizing Spatial Interaction," (May, 1981).

No. 116 Sho-Ichiro Kusumoto, Kanemi Ban, Hajime Wago and Kazumi Asako,
"Rational Savings, Price Expectation and Money Supply in a Growing Economy," (July, 1981).

No. 117 Sho-Ichiro Kusumoto,
"On the Equilibrium Concepts in a General Equilibrium Theory with Public Goods and Taxes II -- "Surplus" Maximum," (June, 1981).

No. 118 Hajime Eto,
"Decision-Theoretical Foundations of the Validities of Technology Forecasting Methods," (June, 1981).

No. 119 Hiroyuki Odagiri,
"Internal Promotion, Intrafirm Wage Structure and Corporate Growth," (July, 1981).

No. 120 Hajime Eto and Kyoko Makino,
"The Validity of the Simon's Firm-Size Model and its Revision," (August, 1981).

No. 121 Satoru Fujishige,
"Structures of Polytopes Determined by Submodular Functions on Crossing Families," (August, 1981).

No. 122 Hajime Eto,
"Epistemologico-Logical Approach to the Validity of Model in View of the Fuzzy System Model," (August, 1981).

No. 123 Hiroyuki Odagiri,
"R & D Expenditures, Royalty Payments, and Sales Growth in Japanese Manufacturing Corporations," (August, 1981).

No. 124 Noboru Sakashita,
"Evalution of Regional Development Policy-An Alternative Approach," (September, 1981).

No. 125 Takao Fukuchi and Noriyoshi Oguchi,
"A Generalization of Keynesian Macromodel; the STIC Model," (September, 1981).

No. 126 Kazumi Asako,
"The Penrose Effect and the Long-Run Equilibrium of a Monetary Optimizing Model: Superneutrality and Nonexistence," (September, 1981).

No. 127 Isao Ohashi,
"Optimal Properties of Wage and Layoff Policies and the Impact of Trade Unionism," (August, 1981).

No. 128 Kanemi Ban,
"A Macroeconometric Model of Japan with Rational Expectations," (September, 1981).

No. 129 Yoshitsugu Yamamoto and Kiyoshi Murata,
"A New Variable Dimension Algorithm : Extension for Separable Mappings, Geometric Interpretation and Some Applications," (October, 1981).

No. 130 Hirotaka Sakasegawa,
"Numerical Analysis on Tandem Queueing System with Blocking," (November, 1981).

No. 131 Ikuo Kabashima,
"Supportive Participatory Model of Development -Political Participation and Income Distribution in Growing Economies- ," (December, 1981).

No. 132 Seizo Ikuta,
"Multistage Decision Process with Random Observations and its Applications," (December, 1981).

No. 133 Yuji Kubo,
"Interindustry Linkages and Industrial Development," (December, 1981).

No. 134 Kazumi Asako and Ryuhei Wakasugi,
"Government Capital, Income Distribution, and Optimal Taxation," (January, 1982).

No. 135 Satoru Fujishige,
"The Theory of Semimodular Programs," (January, 1982).