No. 149 (82-16)

A Note on Submodular Functions On Distributive Lattices

by

Satoru FUJISHIGE

and

Nobuaki TOMIZAWA

April 1982

A NOTE ON SUBMODULAR FUNCTIONS ON DISTRIBUTIVE LATTICES

Satoru FUJISHIGE

Institute of Socio-Economic Planning
University of Tsukuba
Sakura, Ibaraki, Japan 305

and

Nobuaki TOMIZAWA

Graduate School of Science and Engineering
Tokyo Institute of Technology
Nagatsuta, Midori-ku, Yokohama, Japan 227

Abstract — Let $\mathcal{F}$ be a distributive lattice formed by subsets of a finite set E with $\emptyset, E \in \mathcal{F}$ and let R be the set of reals. Also let f be a submodular function from $\mathcal{F}$ into R with $f(\emptyset) = 0$. We determine the set of extreme points of the base polytope

$$\mathrm{B}(f) = \{x \mid x \in R^E, x(D) \leq f(D) \, (D \in \mathcal{F}), x(E) = f(E)\}$$

and give upper and lower bounds of f which can be obtained in polynomial time under mild assumptions.

1. Definitions

Let E be a finite set, $\mathcal{F}$ be a distributive lattice formed by subsets of E with $\emptyset, E \in \mathcal{F}$ and R be the set of reals. Also let f be a submodular function from $\mathcal{F}$ into R, i.e.,

$$f(D_1) + f(D_2) \geqq f(D_1 \cup D_2) + f(D_1 \cap D_2) \tag{1.1}$$

for any $D_1, D_2 \in \mathcal{F}$, and suppose $f(\emptyset) = 0$. Let us define a polytope $B(f)$ by

$$B(f) = \{x \mid x \in R^E, x(D) \leqq f(D)\ (D \in \mathcal{F}), x(E) = f(E)\}, \tag{1.2}$$

where for $D \in \mathcal{F}$ and $x = (x(e) : e \in E) \in R^E$

$$x(D) = \sum_{e \in D} x(e). \tag{1.3}$$

We call the pair $(\mathcal{F}, f)$ a <u>submodular system</u> and the polytope $B(f)$ the <u>base polytope</u> associated with the submodular system $(\mathcal{F}, f)$.

We shall determine the set of extreme points of the base polytope $B(f)$ and give upper and lower bounds of f which can be computed in polynomial time under mild assumptions.

2. Representation of Distributive Lattices

For a finite partially ordered set $\mathcal{P} = (X, \preccurlyeq)$, an ordered pair (W^+, W^-) of subsets W^+ and W^- of X with $W^+ \cap W^- = \emptyset$ and $W^+ \cup W^- = X$ is called a <u>monotone dissection</u> of $\mathcal{P}$ if for every $x^+ \in W^+$ and $x^- \in W^-$ we do not have $x^+ \preccurlyeq x^-$. Here, note that W^+ or W^- may be empty.

The following representation theorem for distributive lattices is classical and may be well known (see [1]).

<u>Theorem 2.1</u>: For any distributive lattice F formed by subsets of a finite set E with $\emptyset, E \in F$, there exists a unique partially ordered set $P = (X, \preccurlyeq)$ such that

(i) X is a partition $\{A_1, A_2, \cdots, A_n\}$ of E and

(ii) $D \in F$ if and only if

$$D = \bigcup\{A_i \mid A_i \in W^-\} \tag{2.1}$$

for some monotone dissection (W^+, W^-) of P.

Conversely, for any partially ordered set $P = (X, \preccurlyeq)$ with X being a partition $\{A_1, A_2, \cdots, A_n\}$ of E, the set F of all the subsets D of E which are expressed as (2.1) for monotone dissections (W^+, W^-) of P is a distributive lattice with respect to set inclusion with $\emptyset, E \in F$.

Given a distributive lattice F, the partially ordered set $P = (X, \preccurlyeq)$ in Theorem 2.1 is determined as follows. For each $e \in E$, let $S(e)$ be the unique minimal element in F with $e \in S(e)$, i.e.,

$$S(e) = \bigcap\{D \mid e \in D \in F\}. \tag{2.2}$$

Define a graph $G = (E, A^*)$ with the vertex set E and the arc set A^* by

$$A^* = \{(e_1, e_2) \mid e_1 \in E, e_2 \in S(e_1)\}. \tag{2.3}$$

The decomposition of G into strongly connected components yields a partition of the vertex set E and a partial order on the partition in a natural way which define the required partially ordered set $P = (X, \preccurlyeq)$.

Without loss of generality we assume throughout the present paper that

"each $A_i \in X$ of the partially ordered set $P = (X, \preccurlyeq)$ has cardinality one" (2.4)

and we express P by $(E, \preccurlyeq)$ instead of $(X, \preccurlyeq)$ with $X = \{\{e\} \mid e \in E\}$.

It should be noted that because of this assumption both $S(e)$ and $S(e) - \{e\}$ belong to F for $S(e)$ $(e \in E)$ defined by (2.2) and that, for any integer i such that $0 \leqq i \leqq |E|$, there exists a set $D \in F$ with $|D| = i$.

3. Extreme Points of the Base Polytope

First, we show the following lemma.

Lemma 3.1: Let

$$A_0 = \emptyset \subsetneqq A_1 \subsetneqq \cdots \subsetneqq A_n = E \qquad (3.1)$$

be a maximal chain in the distributive lattice F. (Note that by the assumption (2.4) $|A_i - A_{i-1}| = 1$ $(1 \leqq i \leqq n)$ and $n = |E|$.) Also define $e_i \in E$ $(1 \leqq i \leqq n)$ by

$$\{e_i\} = A_i - A_{i-1}. \qquad (3.2)$$

Then for a vector $x^* = (x^*(e): e \in E)$ defined by

$$x^*(e_i) = f(A_i) - f(A_{i-1}) \qquad (1 \leqq i \leqq n) \tag{3.3}$$

we have

$$x^*(D) \leqq f(D) \qquad (D \in F), \tag{3.4}$$

$$x^*(E) = f(E), \tag{3.5}$$

i.e., $x^* \in B(f)$.

(Proof) The inequality (3.4) with $D = \emptyset$ and equation (3.5) are trivial.

Suppose that (3.4) is valid for any $D \in F$ with $|D| \leqq k$ for some k such that $0 \leqq k < n$ $(= |E|)$. For any $D^* \in F$ with $|D^*| = k+1$ let e^* be an element of D^* such that

$$\{e^*\} = D^* - A_{i^*-1} \tag{3.6}$$

and

$$A_{i^*} \supseteqq D^* \tag{3.7}$$

for some i^* $(1 \leqq i^* \leqq n)$. Then we have $D^* - \{e^*\} \in F$ and it follows from (3.6) and (3.7) and from the submodularity of f that

$$\begin{aligned} x^*(D^*) &= x^*(e^*) + x^*(D^* - \{e^*\}) \\ &\leqq x^*(e^*) + f(D^* - \{e^*\}) \\ &= f(A_{i^*}) - f(A_{i^*-1}) + f(D^* - \{e^*\}) \\ &\leqq f(D^*). \end{aligned}$$

The lemma follows by induction. Q.E.D.

From Lemma 3.1 we see that the base polytope $B(f)$ is nonempty for any submodular function f.

For any weight vector $w \in R^E$ let us consider the problem:

$$P_w : \text{ Minimize } \sum_{e \in E} w(e)x(e)$$

$$\text{subject to } x \in B(f). \tag{3.8}$$

Suppose that the distinct values of $w(e)$ $(e \in E)$ are given by

$$w_1 < w_2 < \cdots < w_p \tag{3.9}$$

and define

$$S_i = \{e \mid e \in E, w(e) \leqq w_i\} \quad (i=1,2,\cdots,p). \tag{3.10}$$

Lemma 3.2: The problem P_w has a finite optimal solution if and only if for each set S_i $(i=1,2,\cdots,p)$ defined by (3.10) the ordered pair $(E - S_i, S_i)$ is a monotone dissection of $(E, \preccurlyeq)$ which represents the distributive lattice F.

(Proof) The "if" part: By the assumption there exists a maximal chain

$$A_0 = \emptyset \subsetneqq A_1 \subsetneqq \cdots \subsetneqq A_n = E \tag{3.11}$$

in the distributive lattice F such that S_i $(i=1,2,\cdots,p)$ are included in (3.11). Let $x^* \in R^E$ be a vector defined by (3.11), (3.2) and (3.3). Then from Lemma 3.1 we have

$$x^* \in B(f). \tag{3.12}$$

Furthermore, for any vector $y \in B(f)$ we have from (3.3), (3.9) and (3.10)

$$\sum_{e \in E} w(e)y(e) - \sum_{e \in E} w(e)x^*(e)$$
$$= \sum_{i=1}^{p} \sum_{e \in S_i - S_{i-1}} w_i(y(e) - x^*(e))$$

$$= \sum_{i=1}^{p} \left\{ \sum_{e \in S_i} w_i(y(e) - x^*(e)) - \sum_{e \in S_{i-1}} w_i(y(e) - x^*(e)) \right\}$$

$$= \sum_{i=1}^{p-1} (w_{i+1} - w_i) \sum_{e \in S_i} (x^*(e) - y(e)) + \sum_{e \in S_p} w_p(y(e) - x^*(e))$$

$$= \sum_{i=1}^{p-1} (w_{i+1} - w_i)(f(S_i) - y(S_i))$$

$$\geqq 0, \tag{3.13}$$

where $S_0 = \emptyset$ (and $S_p = E$).

Therefore, x^* is an optimal solution of P_w.

The "only if" part: Let x^* be an optimal solution of P_w. If for any S_k $(1 \leqq k \leqq p)$ $(E - S_k, S_k)$ is not a monotone dissection of $(E, \preccurlyeq)$, then there is a pair (e_1, e_2) such that $e_1 \preccurlyeq e_2$, $e_1 \in E - S_k$ and $e_2 \in S_k$. Since for every $D \in F$ if $e_2 \in D$ then $e_1 \in D$, for any $d > 0$ we have

$$y \equiv x + d\chi_{e_2} - d\chi_{e_1} \in \mathrm{B}(f), \tag{3.14}$$

where, for $e \in E$, $\chi_e \in R^E$ and

$$\chi_e(e') = \begin{cases} 1 & (e' = e) \\ 0 & (e' \in E - \{e\}). \end{cases} \tag{3.15}$$

Consequently,

$$\sum_{e \in E} w(e)y(e) - \sum_{e \in E} w(e)x^*(e)$$

$$= (w(e_2) - w(e_1))d$$

$$< 0. \tag{3.16}$$

This contradicts the optimality of x^*. Therefore, $(E - S_k, S_k)$ must be a monotone dissection of $(E, \preccurlyeq)$. Q.E.D.

The proof of the "if" part of Lemma 3.2 is a direct adaptation of a proof of the validity of the greedy algorithm for submodular functions on Boolean lattices 2^E [5].

In the proof of Lemma 3.2 we have already shown the following.

Corollary 3.3: For any weight vector $w \in R^E$, if the problem P_w has a finite optimal solution, an optimal solution x^* is given by

$$x^*(e_i) = f(A_i) - f(A_{i-1}) \qquad (i=1,2,\cdots,n), \qquad (3.17)$$

where

$$A_0 = \emptyset \subsetneqq A_1 \subsetneqq \cdots \subsetneqq A_n = E \qquad (3.18)$$

is a maximal chain in F with

$$\{e_i\} = A_i - A_{i-1} \qquad (i=1,2,\cdots,n) \qquad (3.19)$$

and

$$w(e_1) \leqq w(e_2) \leqq \cdots \leqq w(e_n). \qquad (3.20)$$

Corollary 3.3 provides an algorithm for solving the problem P_w which is an extension of the socalled "greedy algorithm" for (poly-)matroids [3].

Theorem 3.4: The extreme points of $B(f)$ are exactly those which are given by (3.17) - (3.19), each corresponds to a maximal chain (3.18) chosen from F.

(Proof) Because of Corollary 3.3 we have only to show that for any vector x^* given by (3.17) - (3.19) there exists a weight vector $w \in R^E$ such that x^* is a unique optimal solution of the problem P_w. For such a vector x^*, let us choose a weight vector $w \in R^E$ such that

$$w(e_1) < w(e_2) < \cdots < w(e_n). \tag{3.21}$$

Then x^* is an optimal solution of P_w due to Corollary 3.3. Moreover, x^* is a unique optimal solution because for any optimal solution y of P_w we have, similarly as (3.13),

$$\begin{aligned} 0 &= \sum_{e \in E} w(e)y(e) - \sum_{e \in E} w(e)x^*(e) \\ &= \sum_{i=1}^{n-1} (w(e_{i+1}) - w(e_i))(f(A_i) - y(A_i)) \\ &\geqq 0, \end{aligned} \tag{3.22}$$

where $A_i = \{e_1, e_2, \cdots, e_i\}$ $(i=1,2,\cdots,n)$. From (3.21) and (3.22),

$$x(A_i) = f(A_i) = y(A_i) \qquad (i=1,2,\cdots,n),$$

i.e.,

$$x(e) = y(e) \quad (e \in E).$$

This concludes the proof of the theorem. Q.E.D.

Theorem 3.4 is a generalization of the extreme point theorem for (poly-)matroid polytopes by J. Edmonds [2].

4. Upper and Lower Bounds of Submodular Functions

We need some lemma to obtain an upper bound of f.

Lemma 4.1: For a vector $\overline{x} = (\overline{x}(e) : e \in E)$ defined by

$$\overline{x}(e) = f(S(e)) - f(S(e) - \{e\}) \tag{4.1}$$

we have

$$\overline{x}(D) \geqq f(D) \tag{4.2}$$

for any $D \in \mathcal{F}$, where $S(e)$ $(e \in E)$ are defined by (2.2).

(Proof) The inequality (4.2) is trivial for $D = \emptyset$.

Suppose that, for some integer k such that $0 \leqq k < |E|$, (4.2) is valid for any $D \in \mathcal{F}$ with $|D| \leqq k$. Then for any $D^* \in \mathcal{F}$ with $|D^*| = k+1$ let e^* be a maximal element of D^* in $\mathcal{P} = (E, \preccurlyeq)$. By the assumption and the submodularity of f we have

$$\begin{aligned}\overline{x}(D^*) &= \overline{x}(e^*) + \overline{x}(D^* - \{e^*\}) \\ &\leqq \overline{x}(e^*) + f(D^* - \{e^*\}) \\ &= f(S(e^*)) - f(S(e^*) - \{e^*\}) + f(D^* - \{e^*\}) \\ &\leqq f(D^*),\end{aligned}$$

where note that $D^* - \{e^*\} \in \mathcal{F}$ and $S(e^*) \subseteqq D^*$.

Therefore, the lemma follows by induction. Q.E.D.

From Lemma 4.1 we have an upper bound $\overline{B}$ of f given by

$$\overline{B} = \sum\{\overline{x}(e) \mid e \in E, \overline{x}(e) > 0\}. \tag{4.3}$$

Furthermore, a lower bound of f is given as follows. Let x^* be an extreme point of $B(f)$. Then

$$\underline{B} = \sum\{x^*(e) \mid e \in E, x^*(e) < 0\} \tag{4.4}$$

is a lower bound of f since

$$\underline{B} \leqq x^*(D) \leqq f(D) \tag{4.5}$$

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

The upper and lower bounds $\overline{B}$ and $\underline{B}$ given by (4.3) and (4.4), respectively, can be obtained in polynomial time with respect to $|E|$ if we assume that the following two operations are carried out in unit time:

(1) to evaluate $f(D)$ for each $D \varepsilon F$;

(2) to discern whether or not there is a set $D \varepsilon F$ such that $e_1 \varepsilon D$ and $e_2 \notin D$ for each $e_1, e_2 \varepsilon E$.

It should be noted that the Hasse diagram of the partially ordered set $P = (E, \preccurlyeq)$ can be obtained in polynomial time when operation (2) is carried out in unit time.

It was shown in [4] that when f is an integer-valued submodular function the minimization of f can be performed in time polynomially bounded by $|E|$ and $\log B$ under the assumptions that operations (1) and (2) are carried out in unit time and that an integral upper bound B for $|f(D)|$ $(D \varepsilon F)$ is previously known. We see that the latter assumption is not necessary for the polynomial-time solvability of minimizing submodualr functions.

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 Calgary International Conference on Combinatorial Structures and Their Applications (Gordon and Breach, 1970), 69-87.

[3] J. Edmonds, Matroids and the greedy algorithm, Mathematical Programming 1 (1971), 127-136.

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

[5] N. Tomizawa, Theory of hyperspace (III) — Maximum deficiency = minimum residue theorem and its applications, in Proceedings of the Research Meeting on Circuits and Systems (The Institute of Electronics and Communication Engineers of Japan, September 1980), CAS 80-74, 41-46 (in Japanese).