Discrete Methods in Informatics January 16, 2006

Lecture 10 (Submodular function)

Lecturer: *Satoru Iwata* *Scribe*: *Masaru Iwasa and Yukie Nagai*

Submodular functions are the functions that frequently appear in connection with many combinatorial optimization problems. For instance, cut capacity functions of networks and rank functions of matroids are submodular functions. Learning about the properties of them helps our general- and individual-based understanding of associated problems, so submodular functions are interesting.

1 Submodular functions and base polyhedra

Let V be a finite set and $n = |V|$. If a set function $f : 2^V \to \mathbb{R}$ satisfies

$$f(X) + f(Y) \geq f(X \cap Y) + f(X \cup Y)$$

for all $X, Y \subseteq V$, then f is referred to as a *submodular function*. Without loss of generality, we can assume $f(\emptyset) = 0$.

We introduce a space $\mathbb{R}^V = \{x \mid x : V \to \mathbb{R}\}$ as a framework for considering submodular functions. For a vector $x \in \mathbb{R}^V$ and each $v \in V$, $x(v)$ denotes the component of x associated with v. For a subset $Y \subseteq V$, define $x(Y) = \sum_{v \in Y} x(v)$. Then x satisfies $x(\emptyset) = 0$ and can be identified with a function consistently satisfying

$$x(X) + x(Y) = x(X \cap Y) + x(X \cup Y). \tag{1}$$

Let a set function $f : 2^V \to \mathbb{R}$ be a submodular function satisfying $f(\emptyset) = 0$. Associated with f, the *submodular polyhedron* $\mathrm{P}(f)$ and the *base polyhedron* $\mathrm{B}(f)$ are defined as

$$\mathrm{P}(f) = \{x \mid x \in \mathbb{R}^V,\ \forall Y \subseteq V,\ x(Y) \leq f(Y)\},$$
$$\mathrm{B}(f) = \{x \mid x \in \mathrm{P}(f),\ x(V) = f(V)\}.$$

In the case where $|V| = 2$, the submodular polyhedron $\mathrm{P}(f)$ and the base polyhedron $\mathrm{B}(f)$ are represented as in Figure 1.

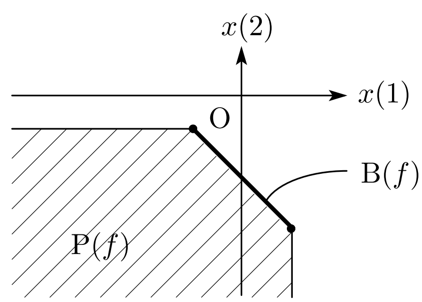

Figure 1. The submodular polyhedron $\mathrm{P}(f)$ and the base polyhedron $\mathrm{B}(f)$.

Proposition 1. *For all $Y, Z \subseteq V$, if a vector $x \in \mathrm{P}(f)$ satisfies $x(Y) = f(Y)$ and $x(Z) = f(Z)$, then $x(Y \cap Z) = f(Y \cap Z)$ and $x(Y \cup Z) = f(Y \cup Z)$.*

Proof. By equation (1) and the submodularity of f,

$$\begin{aligned} x(Y \cap Z) + x(Y \cup Z) &= x(Y) + x(Z) \\ &= f(Y) + f(Z) \\ &\geq f(Y \cap Z) + f(Y \cup Z). \end{aligned} \tag{2}$$

And the vector $x \in \mathrm{P}(f)$ satisfies

$$x(Y \cap Z) \leq f(Y \cap Z), \tag{3}$$
$$x(Y \cup Z) \leq f(Y \cup Z). \tag{4}$$

To hold equation (2), equations (3) and (4) must hold with equalities. Hence, $x(Y \cap Z) = f(Y \cap Z)$ and $x(Y \cup Z) = f(Y \cup Z)$. □

We denote $x \leq y$ when two vectors $x, y \in \mathbb{R}^V$ satisfy $x(v) \leq y(v)$ for every $v \in V$. If that $x \leq y$ for any vector $y \in \mathrm{P}(f)$ implies $x = y$, then a vector $x \in \mathrm{P}(f)$ is called a *maximal vector*.

Proposition 2. *The necessary and sufficient condition for a vector $x \in \mathrm{P}(f)$ to be a maximal vector of $\mathrm{P}(f)$ is that $x \in \mathrm{B}(f)$.*

Proof. It is trivial that x is a maximal vector of $\mathrm{P}(f)$ when $x \in \mathrm{B}(f)$. Now let us show that $x \in \mathrm{B}(f)$ if x is a maximal vector of $\mathrm{P}(f)$.

When x is a maximal vector of $\mathrm{P}(f)$, subset $Y_v \subseteq V$ exists such that $v \in Y_v$ for all $v \in V$, and $x(Y_v) = f(Y_v)$ holds. It is because $V = \bigcup_{v \in V} Y_v$, using Proposition 1, $x(V) = f(V)$ holds. Hence, $x \in \mathrm{B}(f)$. □

On a submodular polyhedron $\mathrm{P}(f)$, a linear programming problem with a nonnegative weight vector $p \in \mathbb{R}^V$ is formulated as follows:

$$\begin{aligned} &\text{maximize} && \sum_{v \in V} p(v)x(v) \\ &\text{subject to} && x \in \mathrm{P}(f). \end{aligned}$$

The optimum solution x of this problem is referred to as the *p-maximal base* and $x \in \mathrm{B}(f)$. In the above formulation, constraints are not represented explicitly, but this problem has $(2^n - 1)$ constraints. As n grows large (for example even about $n = 100$), it is difficult to solve this problem in practical time. However, applying the submodularity of f helps to solve it efficiently.

Let distinct values of each component $p(v)$ be $p_1 > p_2 > \cdots > p_k$ and $U_i = \{v \mid p(v) \geq p_i\}$. U_i is represented as in Figure 2.

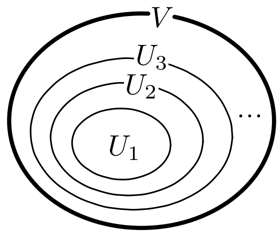

Figure 2. $U_i = \{v \mid p(v) \geq p_i\}$.

The next proposition characterizes the p-maximal base.

Proposition 3. *The necessary and sufficient condition for a vector $x \in \mathrm{B}(f)$ to be a p-maximal base is that for all i, $x(U_i) = f(U_i)$.*

Proof. In the dual problem:

$$\begin{aligned} &\text{minimize} && \sum_{Y \subseteq V} \xi(Y) f(Y) \\ &\text{subject to} && \sum_{v \in Y \subseteq V} \xi(Y) = p(v) && (v \in V), \\ & && \xi(Y) \geq 0 && (Y \subseteq V,\ Y \neq V), \end{aligned}$$

let $\xi(V) = p_k$ for $i = k$ and $\xi(U_i) = p_i - p_{i+1}$ for $i = 1, \ldots, k-1$. Define the other $Y \subseteq V$ as $\xi(Y) = 0$. The variable ξ obtained in this way satisfies the constraints of the dual problem and is a feasible solution. If a vector $x \in \mathrm{B}(f)$ satisfies $x(U_i) = f(U_i)$ for all i, then x and ξ satisfy the complementarity conditions and they are both the optimal solutions. In fact, when the complementarity conditions about ξ:

$$\text{for any } Y \subseteq V, \text{ if } \xi(Y) > 0 \text{ then } x(Y) = f(Y),$$

is satisfied, the following equation holds:

$$\begin{aligned} \sum_{Y \subseteq V} \xi(Y) f(Y) &= p_k f(V) + \sum_{i=1}^{k-1} (p_i - p_{i+1}) f(U_i) \\ &= p_k x(V) + \sum_{i=1}^{k-1} (p_i - p_{i+1}) x(U_i) = \sum_{v \in V} p(v) x(v). \end{aligned}$$

The existence of such a vector $x \in \mathrm{B}(f)$ is shown by the submodularity of f.

On the other hand, $x(U_i) = f(U_i)$ holds because arbitrary p-maximal base x satisfies the complementarity conditions about ξ. $\square$

2 The relationship between submodular functions and convex functions

In this section, we state the proposition proved by L. Lovász that reveals the relationship between submodular functions and convex functions.

Let V be a finite set and $n = |V|$. Then 2^V corresponds to the vertices of a n-dimensional hypercube $Q = \{p \in \mathbb{R}^V \mid 0 \leq p(v) \leq 1\}$. Then we assume for any set function $f : 2^V \to \mathbb{R}$ that satisfies $f(\emptyset) = 0$, the functional value of f is determined at each vertex. When we consider deciding the values of f at the other points, it is done with triangulation and simplicial subdivision.

An n-dimensional hypercube can be decomposed to $n!$ congruent n-simplexes. And values of f at non-vertex points of each n-simplex are linearly interpolated with the values of vertices. The values of f obtained as stated above are continuous in n-dimensional hypercube.

For example, a 3-dimensional cube can be decomposed to $3!$ $(= 6)$ congruent tetrahedrons. In the 3-dimensional cube in Figure 3, of these 6 tetrahedra, the one $P = \{p \in Q \mid p(1) \geq p(2) \geq p(3)\}$ with vertices corresponding to the sets $\emptyset$, $\{1\}$, $\{1,2\}$, and $\{1,2,3\}$ is drawn with heavy lines.

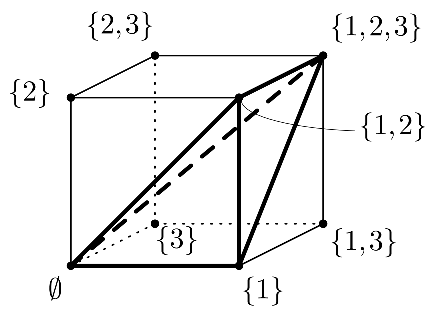

Figure 3. A 3-dimensional cube.

For a non-negative vector $p \in [0,1]^V$, let distinct values of each component $p(v)$ be $p_1 > p_2 > \cdots > p_k$ and $U_i = \{v \mid p(v) \geq p_i\}$. Then the sequence of positive real numbers $\lambda_1, \lambda_2, \ldots, \lambda_{k-1}$ exists uniquely and is represented as

$$p = \sum_{i=1}^{k} \lambda_i \chi_{U_i}.$$

For any set function $f : 2^V \to \mathbb{R}$ satisfying $f(\emptyset) = 0$, determine a function $\hat{f} : \mathbb{R}^V_+ \to \mathbb{R}$ as

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

This definition corresponds to linear interpolation of values of non-vertex points in n-dimensional hypercube. Note that for any $\alpha > 0$, $\hat{f}(\alpha p) = \alpha \hat{f}(p)$ holds.

The next proposition about the relationship between submodular functions and convex functions is told by L. Lovász in 1983.

Proposition 4. *The necessary and sufficient condition for a function f to be submodular is that $\hat{f}$ is convex.*

Proof. When a function f is submodular, by Proposition 3, for any $p \in \mathbb{R}^V$,

$$\hat{f}(p) = \max\left\{ \sum_{v \in V} p(v)x(v) \;\middle|\; x \in \mathrm{P}(f) \right\}$$

holds. Hence, $\hat{f}$ is a convex function.

Note that $\hat{f}(\alpha p) = \alpha \hat{f}(p)$ holds for any $\alpha > 0$, then if a function $\hat{f}$ is convex, for any $Y, Z \subseteq V$

$$\hat{f}(\chi_Y + \chi_Z) = 2 \cdot \hat{f}\left(\frac{\chi_Y + \chi_Z}{2}\right) \leq \hat{f}(\chi_Y) + \hat{f}(\chi_Z) = f(Y) + f(Z)$$

holds. On the other hand, by the definition of $\hat{f}$,

$$\hat{f}(\chi_Y + \chi_Z) = \hat{f}(\chi_{Y \cap Z}) + \hat{f}(\chi_{Y \cup Z}) = \hat{f}(Y \cap Z) + \hat{f}(Y \cup Z).$$

In consequence, f is a submodular function. $\square$

By Proposition 4, submodular functions can be understood as set functions with convexity. It is because minimization of convex functions can be solved in polynomial time, minimization of submodular functions is solvable in polynomial time, too.

References

[1] A. Schrijver: *Combinatorial Optimization.* Springer-Verlag, 2003.