CMPUT 675: Topics in Combinatorics and Optimization **Fall 2016**

Lecture 30 (Nov. 23): Submodular Functions

Lecturer: Zachary Friggstad *Scribe: Arnoosh Golestanian*

Part of this lecture proved the min-max theorem for Matroid Partitioning. The proof was moved to the matroid partitioning notes in Lecture 28.

We moved on to submodular functions and submodular maximization.

30.1 Submodular functions

Definition 1 *For a finite set X, a function $f : 2^X \to R$ is* **submodular** *if $\forall\ A, B \subseteq X, f(A) + f(B) \geq f(A \cup B) + f(A \cap B)$.*

Examples:

- Matroid rank: have been discussed earlier in this course.
- Cardinality function: $f(A) = |A|$
- Cut function: Let A be a subset of vertices in a directed graph, then $f(A) = |\delta^{out}(A)|$
- Coverage function: For $X = \{S_1, S_2, ..., S_n\}$ being a collection of subsets of some ground set Z, the coverage function $f(A) = |\bigcup_{S \in A} S|$, $A \subseteq 2^X$ is submodular.
- Max function: $\max\{z \in A\}$ if $X \subseteq \mathbb{R}$ and $A \subseteq X$.
- $s - t$ cut function on graph $G(V, E)$: Let $X = V - \{s, t\}$, $f(A) = |\delta^{out}(A \cup \{s\})|$.
- Flow function on digraph $G(V, E)$ with capacity $\forall\ e \in E, \mu_e$: Let $r \in V$ and $A \subseteq V - r$, $f(A)$ =maximum flow from r into A in G.

We will assume that $f(A)$ can be evaluated efficiently for any $A \subseteq X$. In this case, say that the function is presented by a **value oracle**. For example, if f is the cut function then for any subset of vertices A we calculate $f(A)$ by checking to see which edges start in A and end outside of A.

Finding a set A with minimum value can be done in polynomial time in the sense that the algorithm uses a polynomial number of queries to the oracle computing f and, apart from these calls, runs in polynomial time. However, it is a fairly complicated algorithm. We will sketch it next lecture. For now, we discuss some simpler maximization algorithms.

The problem we consider is the following,

Definition 2 *Find $A \subseteq X$ such that $f(A)$ has the maximum possible value.*

This is **NP**-hard, as it generalizes the maximum-cut problem. So we will consider some approximation algorithms. The only assumption we will make in our first algorithm is that $f(A) \geq 0$ for each $A \subseteq X$. This holds in many natural problems, such as coverage functions or cut functions.

Our first algorithm is quite oblivious. It picks a set without **ever** querying the value oracle! So it is really more of a structural statement about submodular functions. Still, it can be interpreted as a randomized approximation algorithm.

Approximation Algorithm for Submodular Maximization

1. $A \leftarrow \emptyset$
2. for each $i \in X$
3. add i to A with probability $\frac{1}{2}$
4. return A

Theorem 1 *If $f(A) \geq 0$ for all $A \subseteq X$, then $E[f(A)] \geq \frac{1}{4} \max_{B \subseteq X} f(B)$.*

To prove this theorem, we need to use the following two lemmas. Before doing this, we briefly mention the following easy-to-check fact.

Fact
If $f : 2^X \to \mathbb{R}$ is submodular and $A, B \subseteq X$ are disjoint, then $f' : 2^{X-(B\cup A)} \to \mathbb{R}$ where $f'(T) = f(T \cup A)$ is also submodular.

Lemma 1 $E_{A\subseteq X'}[f'(A)] \geq \frac{1}{2}\left[f'(\emptyset) + f'(X')\right]$ *for any submodular function f' over X'.*

Proof.

We prove this by induction on $|X'| = n$. The base case is when $n = 1$; so, in the algorithm either we choose $\emptyset$ or X' meaning that:

$$E_{A\subseteq X'}[f'(A)] = \frac{1}{2}\left[f'(\emptyset) + f'(X')\right]$$

Now when $n \geq 1$, assume that the claim is true for any subset of set X' and we want show that this is also true for set X'. Let $z \in X'$, then

$$E_{A\subseteq X'}[f'(A')] = \frac{1}{2}\, E_{A'\subseteq X'-z}[f'(A')] + \frac{1}{2}\, E_{A'\subseteq X'-z}[f'(A'+z)]$$

$$\geq \frac{1}{2}\left(\frac{1}{2}[f'(\emptyset) + f'(X'-z)]\right) + \frac{1}{2}\left(\frac{1}{2}[f'(z) + f'(X')]\right) \geq \frac{1}{2}[f'(\emptyset) + f'(X')].$$

The first inequality used the induction hypothesis on the two submodular functions f_1, f_2 over $X - \{z\}$ obtained from f' where $f_1(A') = f'(A')$ and $f_2(A') = f'(A'+z)$ (see the **fact** above). The second inequality follows directly from the definition of submodularity.

∎

Lemma 2 $\forall\, B \subseteq X'$, $E_{A\subseteq X'}[f'(A)] \geq \frac{1}{4}\left[f'(\emptyset) + f'(B) + f'(X'-B) + f'(X')\right]$

Proof.

$$E_{A\subseteq X}[f'(A)] = E_{A'\subseteq B}[E_{\overline{A'}\subseteq X-B}[f'(A'\cup\overline{A'})]] \geq E_{A'\subseteq B}[\frac{1}{2}f'(A') + \frac{1}{2}f'(A'\cup(X'-B))]$$

$$= \frac{1}{2}E_{A'\subseteq B}[f'(A') + \frac{1}{2}E_{A'\subseteq B}[f'(A'\cup(X'-B))]]$$

$$\geq \frac{1}{2}[\frac{1}{2}f'(\emptyset) + \frac{1}{2}f'(B)] + \frac{1}{2}[\frac{1}{2}f'(X'-B) + \frac{1}{2}f'(X')]$$

Here we have used Lemma 1 repeatedly.

■

Proof.

Now we are ready to prove Theorem 1.

By using Lemma 2, letting f' to be f, and assuming that B maximizes f,

$$E_{A\subseteq X}[f(A)] \geq \frac{1}{4}\ [f(\emptyset) + f(B) + f(X-B) + f(X)] \geq \frac{1}{4}f(B)$$

and the reason for the last inequality is because we assumed that f is non negative.

■

Say that f is **symmetric** if $f(A) = f(X - A)$ for each $A \subseteq X$. For example, if f is the cut function of an undirected graph.

Corollary 1 *If f is symmetric and $f(A) \geq 0$ for all $A \subseteq X$, then $E[f(A)] \geq \frac{1}{2}\max_{B\subseteq X} f(B)$.*

Proof. We know
$$E_{A\subseteq X}[f(A)] \geq \frac{1}{4}\ [f(\emptyset) + f(B) + f(X-B) + f(X)] \geq \frac{1}{2}f(B)$$
for any B, in particular for the B that maximizes $f(B)$. ■

The analysis of both the general and the symmetric versions are tight. For example, let $X = \{1,2\}$ and $f(\{1\}) = 1$ and $f(A) = 0$ for all other $A \subseteq X$. For the symmetric case let $f(A) = 1$ if $|A| = 1$ and otherwise $f(A) = 0$.