Topics in Theoretical Computer Science	May 13, 2014

Lecture 12

Lecturer: Moran Feldman *Scribes: Niklas Gehlen, Christos Kalaitzis*

1 Introduction to Set Functions

In this lecture we want to start our discussion on set functions with the focus on submodular functions. Let us begin with defining what a set function is.

Definition 1 *A* set function $f : 2^N \to \mathbb{R}$ *is a function assigning a real value to every subset $S \subseteq N$ of a given ground set N.*

As a motivation we want to start with two very basic set functions. We will use those examples throughout the lecture to illustrate some results. Our first example is the function measuring the size of a cut in a graph $G = (V, E)$ induced by $(S, \bar{S})$. Formally, we define

$$\delta(S) = |\{(u, v) : u \in S, v \in \bar{S}\}|$$

for every subset of nodes $S \subseteq V$. As a second example consider a finite ground set $N = \{S_1, S_2, \ldots, S_n\}$ with $S_i \subset \mathbb{N}$ being finite and the function $c : 2^N \to \mathbb{R}$:

$$c(N') = \left| \bigcup_{S \in N'} S \right|, N' \subseteq N$$

Intuitively, c measures the size of a union. Todays lectures main interest is determining properties of set functions. Using these enables us to identify certain cases that can(or cannot) be solved efficiently. Two simple properties our above examples have in common are nonnegativity and normalization.

Definition 2 *A set function f is nonnegative if for every $S \subseteq N$ we have $f(S) \geq 0$.*

Definition 3 *A set function f is normalized if $f(\emptyset) = 0$.*

Many examples of set functions can be interpreted as a measure of the value we think a collection of items has. If we think of such a value function we do not want the value to decrease as the set increases, which leads to the following property:

Definition 4 *A set function f is monotone if for all $A \subseteq B \subseteq N$ the constraint $f(A) \leq f(B)$ holds.*

An important concept to ponder on are complementary objects. For example we can think of a pair of shoes or a pencil and a sharpener. Having only one of these objects is rather useless whilst adding the second increases the total value substantially. Thus, there is a strong dependency within these pairs, which we want to avoid in general. A first approach to limiting this dependency is subadditivity.

Definition 5 *A set function f is subadditive if for all $A \cap B = \emptyset$ with $A, B \subseteq N$ the value of the union does not exceed the sum of the individual values $f(A) + f(B) \geq f(A \cup B)$.*

Unfortunately, it is easy to show that this property does not fully inhibit complementarity. Take a glimpse at the following example. Let our ground set consist of three elements $N = \{a, b, c\}$. We define f to be:

$$f(S) = \begin{cases} 0 & |S| = 0 \\ 1 & |S| \in \{1, 2\} \\ 2 & |S| = 3 \end{cases}$$

Note that f is subadditive. Nevertheless, by this definition we get:

$$f(\{a, b\}) - f(\{a\}) = 0, f(\{a, c\}) - f(\{a\}) = 0, f(\{a, b, c\}) - f(\{a\}) = 1$$

So in some sense, b and c complement each other as including them both is better than including only one of them.

2 Submodular Functions

Submodularity is a generalization of subadditivity that deals with the last counterexample on non-complementarity.

Definition 6 *A set function f is submodular if $f(A) + f(B) \geq f(A \cup B) + f(A \cap B)$ holds for all $A, B \subseteq N$.*

Definition 7 *Given a set function f, a set $S \subseteq N$ and an element $u \in N$ the* marginal contribution *of u to S w.r.t. f is defined as*

$$f_u(S) = f(S \cup \{u\}) - f(S)$$

Using the latter definition we will give an alternative definition of submodular functions.

Lemma 8 *A set function f is submodular if and only if for all $A \subseteq B \subseteq N$ and each $u \in N \setminus B$ the following holds:*

$$f_u(A) \geq f_u(B)$$

Proof First we want to assume submodularity and show $f_u(A) \geq f_u(B)$. Consider the sets $A \cup \{u\}$ and B. According to submodularity we have:

$$\begin{aligned}
f(A \cup \{u\}) + f(B) &\geq f(A \cup \{u\} \cup B) + f((A \cup \{u\}) \cap B) \\
f(A \cup \{u\}) + f(B) &\geq f(\{u\} \cup B) + f(A) \\
f(A \cup \{u\}) - f(A) &\geq f(\{u\} \cup B) - f(B) \\
f_u(A) &\geq f_u(B)
\end{aligned}$$

Now, let us assume $f_u(A) \geq f_u(B)$ and prove submodularity. Consider two sets $C, D \subseteq N$. Let h be the number of elements in $D \setminus C = \{d_1, d_2, \ldots, d_h\}$. Also let $D_i = \{d_j : 1 \leq j \leq i\}$.

$$\begin{aligned}
f(D) - f(C \cap D) &= \sum_{i=1}^{h} f((C \cap D) \cup D_i) - f((C \cap D) \cup D_{i-1}) \\
&= \sum_{i=1}^{h} f_{d_i}((C \cap D) \cup D_{i-1}) \\
&\geq \sum_{i=1}^{h} f_{d_i}(C \cup D_{i-1}) \\
&= f(C \cup D) - f(C)
\end{aligned}$$

Note that $D_0 = \emptyset$, $D_h = D \setminus C$ and that the first sum is telescopic. Hence, the first equality holds. The other equalities follow by the definition of the marginal contribution. The inequality holds by assumption. In total we get $f(C) + f(D) \geq f(C \cup D) + f(C \cap D)$ which concludes the second direction of the proof. ■

Now we want to analyze our initial two set functions according to the properties that we just defined.

Monotonicity c is monotone as unifying more sets results in a greater union and therefore in a larger value for c. On the other hand, δ is not monotone. This can be seen with the following example:

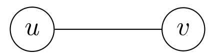

Figure 1: $\delta(\{v\}) = 1, \delta(\{v, u\}) = 0$

Submodularity c is submodular. This can be seen with the marginal contribution of a set $S \in N$:

$$c_S(N') = \left| \bigcup_{T \in N'} T \cup S \right| - \left| \bigcup_{T \in N'} T \right| = \left| S \setminus \bigcup_{T \in N'} T \right|$$

The last term is decreasing in N' which proves submodularity. δ is also submodular. To see this we again want to measure the marginal contribution. Let $E(v, V')$ be the number of edges between some node v and a set of nodes V'.

$$\delta_v(V') = E(v, V \setminus (V' \cup \{v\})) - E(v, V')$$

The minuend is the number of edges additionally cut when adding v which is decreasing in V'. The subtrahend is the number of edges that will not be cut anymore after adding v and is increasing in V'. As before this proves the submodularity.

3 Exercise 1

Given a graph $G = (V, E)$ let $r : 2^E \to \mathbb{R}$ be a set function assigning to every set of edges $E' \subseteq E$ the size of the largest forest included in E'. Prove that r is normalized, monotone and submodular.
Proof The only subforest in $\emptyset$ is itself with respective size 0. Hence r is normalized.
To prove monotonicity let $A \subseteq B \subseteq E$ be some edge sets and let T be a maximal forest in A. We have $T \subseteq A \subseteq B$ so T is a forest in B. Therefore $|T| = r(A)$ is a lower bound on $r(B)$.
Finally, we want to prove submodularity. As before we will do so by bounding the marginal contribution of one element. Let $A \subseteq B \subseteq E$ and $e \in E \setminus B$. Note that the marginal contribution can only take values in $\{0, 1\}$ as adding one edge can increase the largest forest size by at most one. Thus we only need to prove $r_e(B) = 1 \implies r_e(A) = 1$.
Each connected component adds number of nodes minus one edges to the maximum forest. This leads to the following definition of r:

$$r(B) = |V| - \#\text{connected components in } (V, B)$$

Hence e has a marginal contribution of 1 only if it connects two components in (V, B).

$$\begin{aligned} f_e(B) = 1 &\implies e \text{ connects two components in } (V, B) \\ &\implies e \text{ connects two components in } (V, A) \\ &\implies r_e(A) = 1 \end{aligned}$$

■

4 Submodular function minimization

Next, we will start concerning ourselves with some algorithmic problems concerning submodular functions. Since we will want that our algorithms are efficient, a first issue is how the examined submodular function is given to us as part of the input. Clearly, providing the value of such a function for every possible subset of the ground set N might require a lot of space(exponential in $|N|$), while nearly every interesting problem concerning submodular functions suddenly becomes trivial, as we can go over all subsets of N in linear time and find the one that satisfies the requirements of the examined problem(e.g. if we are interested in maximizing a submodular function, we can find the subset of N with the maximum value).

Therefore, from now on we will assume that f is given in the form of access to an oracle, i.e. for every $S \subseteq N$ we can in constant time query the value $f(S)$.

The first problem we will examine is called unconstrained submodular function minimization(i.e. USmin), and is the problem of finding the $S \subseteq N$ which minimizes f. As we will see, this is a rare case of a problem of substantial interest that we can solve exactly in polynomial time. Since for problems we cannot solve exactly we will employ approximation algorithms, for nearly every other interesting problem, we will require non-negativity, since otherwise the concept of approximation ratio is not very meaningful(what approximation ratio would we expect for the problem of maximizing a submodular function whose maximum can be 0?).

The way we will do this is by showing that it is equivalent to finding the global minimum of a certain convex function, a problem we can solve using the ellipsoid method in polynomial time.

Definition 9 *Given $f : 2^N \to \mathbb{R}$, the convex closure(cc) f^- of f is a function $f^- : [0,1]^N \to \mathbb{R}$ such that, if $D^-(x)$ for $x \in [0,1]^N$ is a distribution on subsets of N satisfying the marginals of x(i.e. $\Pr_{S\sim D^-(x)}[u \in S] = x_u$) which minimizes*

$$\mathbb{E}_{S\sim D^-(x)}[f(S)]$$

then $f^-(x) = \mathbb{E}_{S\sim D^-(x)}[f(S)]$.

The connection between f and f^- is highighted in the following lemma:

Lemma 10 *Let f^- be the cc of f. Then*

- *f^- is an extension of f, i.e. for all $S \subseteq N$, $f(S) = f^-(\mathbf{1}_S)$, where $\mathbf{1}_S$ is the indicator vector of S.*
- $\min\limits_{S\subseteq N} f(S) = \min\limits_{x\in[0,1]^N} f^-(x)$
- *f^- is convex*
- *Every convex function $g : [0,1]^N \to \mathbb{R}$ for which $f(S) \geq g(\mathbf{1}_S)$ for all $S \subseteq N$ must also obey $f^-(x) \geq g(x)$, for all $x \in [0,1]^N$.*

Proof

- Follows immediately from the fact that there is only one distribution satisfying the marginals of $\mathbf{1}_S$.
- Because of the first property, $\min\limits_{S\subseteq N} f(S) \geq \min\limits_{x\in[0,1]^N} f^-(x)$. On the other hand, since $f^-(x)$ is the average of some values of f, $\min\limits_{S\subseteq N} f(S) \leq f^-(x)$ for all $x \in [0,1]^N$.

- We need to prove that for all $x, y \in [0,1]^N$ $\lambda f^-(x) + (1-\lambda)f^-(y) \geq f^-(\lambda x + (1-\lambda)y)$. Let D' be the probability distribution that with probability λ returns a set sampled according to $D^-(x)$ and otherwise returns a set sampled according to $D^-(y)$. Now notice that

$$\lambda f^-(x) + (1-\lambda)f^-(y) = \lambda \mathbb{E}_{S\sim D^-(x)}[f(S)] + (1-\lambda)\mathbb{E}_{S\sim D^-(y)}[f(S)] = \mathbb{E}_{S\sim D'}[f(S)]$$

Notice that D' satisfies the marginals of $\lambda x + (1-\lambda)y$, which means that D' is a candidate for $D^-(\lambda x + (1-\lambda)y)$ and therefore

$$\mathbb{E}_{S\sim D'}[f(S)] \geq f^-(\lambda x + (1-\lambda)y)$$

- Notice that $x = \sum\limits_{S\subseteq N} \mathbf{1}_S \Pr[S]$. Hence, since g is convex,

$$g(x) \leq \sum_{S\subseteq N} g(\mathbf{1}_S)\Pr[S] \leq \sum_{S\subseteq N} f(S)\Pr[S] \leq f^-(x)$$

■

Let us note that now our problem reduces to evaluating f^-; if we have an algorithm for evaluating f^- in polynomial time, then since it is a convex function we can minimize it in polynomial time, thus returning a minimum for f as well.

Unfortunately, we cannot evaluate the convex closure of general set functions; however, it is possible for submodular functions. In fact, the best known algorithm for minimizing an unconstrained submodular function runs in $O(n^6)$ time and requires $O(n^5)$ oracle queries, where $n = |N|$.

Let us now proceed with describing how we can evaluate f^- if f is submodular.

Definition 11 *Given $x \in [0,1]^N$ and $\lambda \in [0,1]$, let $T_\lambda(x) = \{u \in N : x_u \geq \lambda\}$. Given a set function f, the Lovász extension of f is defined as*

$$\hat{f}(x) = \int_0^1 f(T_\lambda(x))d\lambda$$

In other words, $\hat{f}(x)$ is the expected value of $f(T_\lambda(x))$ when λ is chosen uniformly at random from $[0,1]$. Clearly, $\hat{f}(x)$ can be evaluated, since it can be decomposed into at most n terms, the coefficients of which can be easily computed.

Next, we will have to set up a connection between evaluating $\hat{f}(x)$ and $f^-(x)$ if f is submodular:

Lemma 12 *$f : 2^N \to \mathbb{R}$ is submodular iff $f^-(x) = \hat{f}(x)$ for all $x \in [0,1]^N$.*

Proof We will only prove one direction of the Lemma, the one that is required in order to prove the existence of a polynomial time algorithm for USmin.

Let f be submodular; we will prove that $f^-(x) = \hat{f}(x)$ for all $x \in [0,1]^N$. Fix some $x \in [0,1]^N$ and let D be a probability distribution that is a candidate for $D^-(x)$ which maximizes

$$\mathbb{E}_{S\sim D}[|S|^2]$$

The existence of a D which maximizes $\mathbb{E}_{S\sim D}[|S|^2]$ will not be proved, but follows from a compactness argument.

The first thing we will do is prove that the support of D only contains sets that are either disjoint or included in one another. We will call $S_1, S_2 \subseteq N$ crossing sets if $S_1 \setminus S_2 \neq \emptyset$ and $S_2 \setminus S_1 \neq \emptyset$.

Let's assume for the sake of contradiction that D returns two crossing sets S_1, S_2 with probability p_1, p_2. Now, create a distribution D' which decreases the probability of selecting S_1 and S_2 by $\min\{p_1, p_2\}$

and increases the probability of selecting $S_1 \cup S_2$ and $S_1 \cap S_2$ by $\min\{p_1, p_2\}$. By submodularity, we have that

$$f(S_1) + f(S_2) \geq f(S_1 \cap S_2) + f(S_1 \cup S_2)$$

and hence

$$\mathbb{E}_{S \sim D}[f(S)] \geq \mathbb{E}_{S \sim D'}[f(S)]$$

Let $a = |S_1 \setminus S_2|$, $b = |S_2 \setminus S_1|$ and $c = |S_1 \cap S_2|$. Then

$$|S_1 \cup S_2|^2 + |S_1 \cap S_2|^2 = (a + b + c)^2 + c^2 = |S_1|^2 + |S_2|^2 + 2ab > |S_1|^2 + |S_2|^2$$

which implies that D does not maximize

$$\mathbb{E}_{S \sim D}[|S|^2]$$

Furthermore, if such a D' existed, it would be a candidate for $D^-(x)$ along with D; hence, since D already maximizes

$$\mathbb{E}_{S \sim D}[|S|^2]$$

we have a contradiction, which means that no such D' exists, and hence D(which we remind is a candidate for $D^-(x)$) must not contain any crossing sets in its support.

Finally, it can be proved that among all distributions that are candidates for $D^-(x)$ and contain no crossing sets in their support, the one that agrees with $D^-(x)$ is exactly the one defined by the Lovász extension(intuition: when designing such a distribution, we are constrained to satisfy the marginals, we want to minimize $\mathbb{E}_{S \sim D}[f(S)]$ and we know that since f is submodular the marginal returns are diminishing; therefore, it makes sense that we will pick sets of small cardinality with as low probability as possible, which is exactly what sampling $T_\lambda(x)$ by sampling $\lambda \in_R [0, 1]$ does; notice this is also a good strategy to maximize $\mathbb{E}_{S \sim D}[|S|^2]$). ■

Therefore, we know that $\hat{f}(x)$ and $f^-(x)$ agree for submodular functions, and we know how to evaluate $\hat{f}(x)$; hence, our exposition is concluded.