Algorithms II November 25, 2025

Lecture 20: Submodularity and Minimization

Notes by Ola Svensson

In these notes, we introduce submodular functions, give examples, and explain how they can be minimized in polynomial time via Lovász extension. The notes are based on a mix of the following resources:

- Last year's notes written by Moran Feldman and Justin Ward who were postdocs at EPFL and are now faculty at the Open University and Queens Mary University, respectively.
- The notes
  - `https://www.cs.princeton.edu/courses/archive/fall16/cos521/Lectures/lec16.pdf` by Pravesh Kothari (Princeton) which in turn are based on the notes
  - `http://people.csail.mit.edu/moitra/docs/6854notes13.pdf` by Ankur Moitra (MIT).

1 Introduction to Set Functions and Submodularity

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*.*

Many examples of set functions can be interpreted as a measure of the value we think a collection of items has. We have already seen some examples of set functions "in disguise," when considering weighted optimization problems. For example, in the maximum weight matching problem we have a graph $G = (V, E)$, and a non-negative weight w_e for each edge $e \in E$ and want to find a maximum weight set M of edges from E that form a matching in G. We can think of this as the problem of maximizing the set function:

$$f(M) = \sum_{e \in M} w_e$$

subject to the constraint that M is a matching of G. Here, the value of a collection M is simply the *sum* of the individual value of each of its elements. That is, each element e has some inherent value w_e that does not depend on the other elements we choose in M. In many settings, however, the value of an item may depend on whether we have already selected some other item(s). For example: consider a pair of shoes. The amount you would pay for the left shoe alone is not very large compared to the amount you would pay for both shoes together. In the other direction, suppose you are building a coin (or stamp, or Pokémon) collection. Once you have a coin of a particular type, the value of adding another coin of the same type is not so large anymore. In the rest of the lecture, we consider these types of item interactions in more detail. In particular, submodularity is a generalization of the sort of item dependencies captured in the second example above. Formally, we have the following.

Definition 2 *A set function* f *is submodular if*

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

holds for all $A, B \subseteq N$*.*

In order to get more intuition about submodularity, let's think about the value of a single item may change.

Definition 3 *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*

$$f(u \mid S) = f(S \cup \{u\}) - f(S)\,.$$

If we interpret f as assigning a value to collections of items, then $f(u \mid S)$ is precisely how much extra value we would gain by taking element u when we had already taken all elements in S. Using this definition, we can give an alternative definition of submodular functions.

Lemma 4 *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:*

$$f(u \mid A) \geq f(u \mid B)$$

Proof First suppose that f is submodular and consider two sets $A \subseteq B$. We will show that $f(u \mid A) \geq f(u \mid B)$. Consider the sets $A \cup \{u\}$ and B. According to submodularity we have:

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

where in the second line, we have used that $A \subseteq B$.

Now, let us assume $f(u|A) \geq f(u|B)$ for all u and all $A \subseteq B$. We will show that f must be submodular. Consider any 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 $D_i = \{d_j : 1 \leq j \leq i\}$. Then,

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

Note that $D_0 = \emptyset$, $D_h = D \setminus C$ and that the first and last sums are telescopic. Hence, the first and last equalities hold. The other equalities follow by the definition of the marginal contribution. The inequality holds by assumption, since $(C \cap D) \cup D_{i-1} \subseteq C \cup D_{i-1}$. Rearranging, we get $f(C) + f(D) \geq f(C \cup D) + f(C \cap D)$, as desired. ■

The alternative characterization given by Lemma 4 gives us some better intuition about submodularity. It says that a set function f is submodular if and only if the marginal value of an element u with respect to some set S can only decrease (or stay the same) as more elements are chosen in S. That is, submodularity captures the common economic notion of *diminishing returns*. In some sense, submodularity can be thought of as a discrete analogue of *concavity*, since it implies that the "discrete derivative" $f(e|S)$ is *non-increasing* in S.

1.1 Examples of submodular functions

There are plentiful examples of useful and interesting submodular functions. Here we give some of our favorites. For the first three, we prove submodularity. For the remaining, we leave that as an exercise.

- Our first example is the function measuring the size of a cut in a graph $G = (V, E)$ induced by $(S, V \setminus S)$. Formally, we consider the ground set V and the define the function $\delta : 2^V \to \mathbb{R}$ by

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

  for every subset of nodes $S \subseteq V$. To see that δ is submodular we want to measure the marginal contribution. Let $E(v, T)$ be the number of edges between some node v and a set of nodes T.

  $$\delta(v \mid S) = E(v, V \setminus (S \cup \{v\})) - E(v, S)$$

  Observe that the first term on the right is decreasing in S while the second term is increasing in S. Thus, the entire right hand side is decreasing in S. This proves the submodularity (via the "diminishing returns" definition given by Lemma 4).

- Consider a finite collection of sets $T_1, T_2 \ldots, T_n$, where each $T_i \subseteq \mathbb{N}$ is a finite set. We consider the ground set $N = \{1, 2, \ldots, n\}$ and set function $c : 2^N \to \mathbb{R}$ by:

  $$c(S) = \left| \bigcup_{i \in S} T_i \right|,$$

  for every $S \subseteq N$. Intuitively c measures the number of elements from $\mathbb{N}$ contained in the union of the sets specified by (indices from) S. This kind of function is often referred to as a coverage function. To show that c is submodular we (again) analyze the marginal contribution of a set T_i:

  $$c(i \mid S) = \left| T_i \cup \bigcup_{j \in S} T_j \right| - \left| \bigcup_{j \in S} T_j \right| = \left| T_i \setminus \bigcup_{j \in S} T_j \right|$$

  The last term is decreasing in S which, as in the previous case, proves submodularity.

- Let $\mathcal{M} = (\mathcal{I}, X)$ be a matroid on ground set X. The rank function $r : 2^X \to \mathbb{R}$ of the matroid is defined by $r(A) = \max_{I \in \mathcal{I} \cap A} |I|$ for every $A \subseteq X$. That is, $r(A)$ is the size of a maximal independent containing only the elements in A. We will show that r is submodular. Consider any two subsets A and B of X. Let C be any maximal independent set of $\mathcal{M}$ contained in $A \cap B$. By the matroid augmentation axiom, we can extend C to a maximal independent set D contained in $A \cup B$. Then:

  $$r(A \cup B) + r(A \cap B) = |D| + |C| \leq |D \cap (A \cup B)| + |D \cap (A \cap B)| = |D \cap B| + |D \cap A| \leq r(B) + r(A).$$

  the penultimate equality follows from the inclusion-exclusion principle, and the last inequality follows since both $D \cap B$ and $D \cap A$ must be independent (as they are subsets of an independent set D) and are contained in B and A, respectively.

- When summarizing data the utility function is often submodular. This is for the same reason as in the previous coin collection example. Suppose e.g. you wish to summarize all the photos of animals in Switzerland. It is better to select one goat and one horse than two horses or two goats. This is exactly the diminishing returns property.

- Another example is influence maximization: suppose you wish to select k persons to give free samples to in order to launch your product. You wish to select k persons that have the maximum influence: they tell the most people about the product. As you will see in the exercise session, this can again be modeled as the problem of maximizing a submodular function subject to a cardinality constraint.

- Our final example is from information theory. Let $x_1, x_2, \ldots, x_n$ be discrete random variables. For any $A \subseteq [n]$, let $H(A)$ be the joint entropy of the variables $\{x_i\}_{i \in A}$. In other words, $H(A) = -\sum_a \mathbb{P}\left[\{x_i\}_{i \in A} = a\right] \log\left(\mathbb{P}\left[\{x_i\}_{i \in A} = a\right]\right)$. One can see that H is a submodular function.