Advanced Algorithms: Solution of Problem 5

Comment. By no means your solutions are expected to be as long as the ones I am providing. Mine are long because I describe the discovery process.

Exercise 1

Monotonicity is trivial. For submodularity, note that for any $e \in E$ and $S \subseteq E$, we have $f(S \cup \{e\}) - f(S) = |e \setminus \cup_{e_1 \in S} e_1|$, which immediately gives the desired property.

Exercise 2

Notation. In the solution I sometimes use $f(e_{i_1}, \ldots, e_{i_r})$ to denote $f(\{e_{i_1}, \ldots, e_{i_r}\})$.

The special case

Here is perhaps the simplest non-trivial special case:

- Three elements: $E = \{e_1, e_2, e_3\}$.
- $k = 2$.
- $f : 2^E \to \mathbb{R}_+$ monotone, submodular, and such that $f(e_1, e_2) = f(e_1, e_2, e_3)$, and thus $OPT = \{e_1, e_2\}$.
- Finally, suppose that Greedy first picks e_3, and then picks e_1.

We will prove that Greedy here achieves a 3/4-approximation, i.e., $f(e_1, e_3) \geq \frac{3}{4} f(e_1, e_2)$. The proof will reveal how to tackle the general case.

Even though we have simplified things a great deal by considering this special case, things are still a bit abstract. Let's look at a concrete example of an f to elucidate the situation.

Coverage. Consider e_1, e_2, e_3 to be subsets of a set U, and assume that $e_1 \cup e_2 = U$. The function f here is $f(S) = |\cup_{e \in S} e|$, where $S \subseteq \{e_1, e_2, e_3\}$ (see Figure 1 for an illustration). We know how to prove the approximation ratio here. We start by arguing that the first set that Greedy picks (i.e., e_3) is large. Concretely,

$$|e_3| \geq \max(|e_1|, |e_2|) \geq \frac{|e_1| + |e_2|}{2} \geq \frac{|e_1 \cup e_2|}{2}$$

or equivalently,

$$f(e_3) \geq \max(f(e_1), f(e_2)) \geq \frac{f(e_1) + f(e_2)}{2} \geq \frac{f(e_1, e_2)}{2} \quad (1)$$

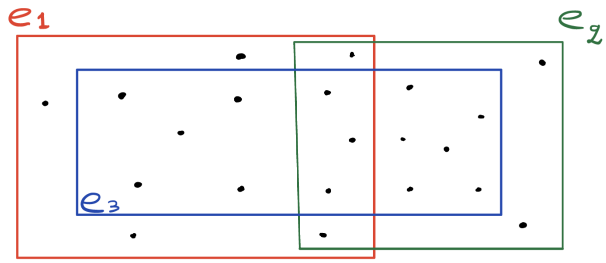

Figure 1: Illustration of the coverage example.

Going back to our special case, it is natural to ask: is 1 true here? It suffices to prove that

$$f(e_1, e_2) \leq f(e_1) + f(e_2) \tag{2}$$

Here is the proof: using submodularity and that $f(\emptyset) \geq 0$, we get

$$f(e_1, e_2) - f(e_1) \leq f(\emptyset \cup \{e_2\}) - f(\emptyset) \leq f(e_2) \tag{3}$$

So, we know that $f(e_3)$ is at least half the optimal: $f(e_3) \geq \frac{f(OPT)}{2}$. Now, we need to analyze the second step of the Greedy: picking e_1. We go back to the Coverage example for inspiration.

Coverage continued. As you know from the analysis of Max-Coverage, to analyze the second step of the Greedy here, we need to use that

$$|e_1 \setminus e_3| + |e_2 \setminus e_3| \geq |(e_1 \cup e_2) \setminus e_3| = |e_1 \cup e_2| - |e_3|$$

The RHS is $f(e_1, e_2) - f(e_3) = f(e_1, e_2, e_3) - f(e_3)$. How to express the LHS using f? Here is the inequality rewritten:

$$(f(e_1, e_3) - f(e_3)) + (f(e_2, e_3) - f(e_3)) \geq f(e_1, e_2, e_3) - f(e_3) \tag{4}$$

We will prove 4 for our special case. First, let's observe that a particular function shows up in 4: the function $g : 2^E \to \mathbb{R}_+$ defined as $g(S) := f(S \cup \{e_3\}) - f(e_3)$, and using g, inequality 4 can be rewritten as $g(e_1) + g(e_2) \geq g(e_1, e_2)$. Exactly the inequality 2, but for g! It is natural to ask the question:

Is g submodular?

Yes, it is! (why?) Thus, the steps in 3 hold for g. Now that we have proven inequality 4, we can show the 3/4-approximation ratio: since the Greedy selected e_1 at the second step, we have $f(e_1, e_3) \geq f(e_2, e_3)$, and thus

$$f(e_1, e_3) - f(e_3) \geq \frac{(f(e_1, e_3) - f(e_3)) + (f(e_2, e_3) - f(e_3))}{2} \geq \frac{f(e_1, e_2, e_3) - f(e_3)}{2} \tag{5}$$

which gives that $f(e_1, e_3) \geq \frac{f(OPT)+f(e_3)}{2}$. We have showed that $f(e_3) \geq \frac{f(OPT)}{2}$, which completes the proof.

The general case

Here $E = \{e_1, e_2, \dots, e_n\}$ and f is monotone and submodular. We first generalize inequality 2:

Claim 1. *Let $g : 2^E \to \mathbb{R}_+$ submodular. Let $S \subseteq E$ such that $S \neq \emptyset$. Then, $g(S) \leq \sum_{e \in S} g(e)$.*

This property is called "subadditivity". Here is the proof:

Proof. Suppose $S = \{e_{i_1}, \dots, e_{i_m}\}$. Using submodularity,

$$g(S) - g(S \setminus \{e_{i_m}\}) \leq g(\emptyset \cup \{e_{i_m}\}) - g(\emptyset)$$

and by rearranging and using that $g(\emptyset) \geq 0$, we have

$$g(S) \leq g(e_{i_m}) + g(S \setminus \{e_{i_m}\})$$

Thus, an induction finishes the proof. □

Now, suppose Greedy chooses the elements $e_{i_1}, \dots, e_{i_k}$ (with this order). Let $OPT \subseteq E$ be the optimal subset of size k. Inequality 5 (of the special case) indicates what inequality we should be aiming for: let $S_j := \{e_{i_1}, \dots, e_{i_j}\}$, i.e., the set collected by Greedy up to step j. We define $S_0 := \emptyset$. We want to show that for all $j \in \{0, 1, \dots, k-1\}$,

$$f(S_j \cup \{e_{i_{j+1}}\}) - f(S_j) \geq \frac{f(S_j \cup OPT) - f(S_j)}{k} \tag{6}$$

Now inequality 4 (of the special case) indicates that to prove 6, we need to first prove that

$$\sum_{e \in OPT \setminus S_j} (f(S_j \cup \{e\}) - f(S_j)) \geq f(S_j \cup OPT) - f(S_j) \tag{7}$$

Observe that 7 indeed implies 6, because of the way Greedy chose $e_{i_{j+1}}$. To prove 7, we generalize the function g of the special case: for any $A \subseteq E$, we define $g_A : 2^E \to \mathbb{R}_+$, given by $g_A(S) := f(A \cup S) - f(A)$. Then, g_A is submodular (why?)[1]. Now, 7 can be rewritten as

$$\sum_{e \in OPT \setminus S_j} g_{S_j}(e) \geq g_{S_j}(OPT \setminus S_j)$$

which is true from Claim 1. Thus, we have proved 6. To show the $(1 - 1/e)$-approximation ratio, we follow the same strategy as in the proof for Max-Coverage, i.e., we consider (for each step) the difference from the optimal value: $\Delta_j := f(OPT) - f(S_j)$. By adding and subtracting $f(OPT)$ in the LHS of 6, we get

$$\Delta_j - \Delta_{j+1} \geq \frac{f(S_j \cup OPT) - f(S_j)}{k} \geq \frac{f(OPT) - f(S_j)}{k} = \frac{\Delta_j}{k}$$

and so $\Delta_{j+1} \leq (1 - 1/k)\Delta_j$ for all $j \in \{0, 1, \dots, k-1\}$, which gives that

$$\begin{aligned} f(OPT) - f(S_k) = \Delta_k &\leq (1 - 1/k)^k \Delta_0 = (1 - 1/k)^k (f(OPT) - f(\emptyset)) \\ &\leq (1 - 1/k)^k f(OPT) \leq \frac{1}{e} f(OPT) \end{aligned}$$

Rearranging finishes the proof.

[1] It is also monotone, but why don't need that.