Exercise Set IX, Advanced Algorithms 2022

These exercises are for your own benefit. Feel free to collaborate and share your answers with other students. Solve as many problems as you can and ask for help if you get stuck for too long. Problems marked * are more difficult but also more fun :).

These problems are taken from various sources at EPFL and on the Internet, too numerous to cite individually.

1 Alice, Bob and Charlie. Suppose that Alice and Bob have two documents d_A and d_B respectively, and Charlie wants to learn about the difference between them. We represent each document by its word frequency vector as follows. We assume that words in d_A and d_B come from some dictionary of size n. Suppose further that the number of distinct words that occur in d_A or d_B but not in both documents is bounded by an integer $k \geq 1$.

Show that Alice and Bob can each send a $O(k \log^2 n)$-bit message to Charlie, from which Charlie can recover the words that occur in d_A or d_B but not in both with probability at least $9/10$. You may assume that Alice, Bob and Charlie have a source of shared random bits, and know the value of n.

Solution: Alice forms a vector $x \in \mathbb{R}^n$ that contains 1s in coordinates corresponding to words that occur in d_A and 0 in other coordinates. Similarly, Bob forms a vector $y \in \mathbb{R}^n$ that contains 1s in coordinates corresponding to words that occur in d_A and 0 in other coordinates.

Let A be a CountSketch matrix with $O(k)$ buckets and $O(\log n)$ repetitions. Alice and Bob both compute Ax and Ay ($O(k \log^2 n)$ bits each) and send these bits to Charlie. Charlie forms $A(x - y) = Ax - Ay$ (here we use the fact that the AMS sketch is a linear map), and recovers the symmetric difference of x and y from the sketch since $x - y$ contains at most k nonzero coordinates, and is therefore exactly reconstructed by CountSketch.

2 Let $f : 2^N \to \mathbb{R}$ be a submodular function. Show that the following functions are also submodular:

- $g(S) = f(S \cup A)$ where A is a fixed set.
- $g(S) = f(S \cap A)$ where A is a fixed set.
- $g(S) = f(N \setminus S)$.

Solution:

- Consider the marginal value of some item $u \in N$. We have

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

for every set S. Let $C \subseteq D$. Then, $C \cup A \subseteq D \cup A$. Hence:

$$g(u \,|\, C) = f(u \,|\, (C \cup A)) \geq f(u \,|\, (D \cup A)) = g(u \,|\, D).$$

- Consider the marginal value of some item $u \in N$. We have

$$g(u \,|\, S) = f((\{u\} \cup S) \cap A) - f(S \cap A) = f((\{u\} \cap A) \cup (S \cap A)) - f(S \cap A)$$

Note that if $u \notin A$ we have $g(u \,|\, S) = 0$ for all S. Now suppose that $u \in A$. Then,

$$g(u \,|\, S) = f(\{u\} \cup (S \cap A)) - f(S \cap A) = f(u \,|\, (S \cap A))$$

for all sets S. Consider any sets $C \subseteq D$. Then, observe that $C \cap A \subseteq D \cap A$, so:

$$g(u \,|\, C) = f(u \,|\, (C \cap A)) \geq f(u \,|\, (D \cap A)) = g(u \,|\, D).$$

- Consider the marginal value of some item $u \in N$. Let $C \subseteq D$. Then, $N \setminus (D \cup \{u\}) \subseteq N \setminus (C \cup \{u\})$, and so:

$$\begin{aligned} g(u \,|\, C) &= f(N \setminus (C \cup \{u\})) - f(N \setminus C) \\ &= -f(u \,|\, (N \setminus (C \cup \{u\}))) \\ &\geq -f(u \,|\, (N \setminus (D \cup \{u\}))) \\ &= f(N \setminus (D \cup \{u\})) - f(N \setminus D) \\ &= g(u \,|\, D). \end{aligned}$$

3 Consider a directed $G = (V, E)$ and define the set function $f : 2^V \to \mathbb{R}$ by

$$f(S) = |\{(u, v) \in E : u \in S, v \notin S\}| \qquad \text{for every } S \subseteq V\,.$$

That is, $f(S)$ equals the number of arcs that exits the set S.

3a Show that f is a (non-monotone) submodular function

Solution: Let $A \subseteq B \subseteq V$ and let $u \in V \setminus B$. We show that $f(u \,|\, A) \geq f(u \,|\, B)$. We have that

$$f(u \,|\, A) = f(\{u\} \cup A) - f(A) = (\text{\# of arcs from } u \text{ to } V \setminus A\ ) - (\text{\# of arcs from } A \text{ to } u\ ),$$

and similarly,

$$f(u \,|\, B) = (\text{\# of arcs from } u \text{ to } V \setminus B\ ) - (\text{\# of arcs from } B \text{ to } u\ ).$$

But, because $A \subseteq B$, we have that

$$(\text{\# of arcs from } u \text{ to } V \setminus B\ ) \leq (\text{\# of arcs from } u \text{ to } V \setminus A\ )$$

and

$$(\text{\# of arcs from } B \text{ to } u\ ) \geq (\text{\# of arcs from } A \text{ to } u\ ),$$

which implies

$$\begin{aligned} f(u \,|\, B) &= (\text{\# of arcs from } u \text{ to } V \setminus B\ ) - (\text{\# of arcs from } B \text{ to } u\ ) \\ &\leq (\text{\# of arcs from } u \text{ to } V \setminus A\ ) - (\text{\# of arcs from } A \text{ to } u\ ) \\ &= f(u \,|\, A). \end{aligned}$$

Hence f is submodular.

Notice that f is not monotone unless $E = \emptyset$. It is easy to see that both $f(\emptyset) = 0$ and $f(V) = 0$, but if there is at least one arc $(u, v) \in E$ then $f(\{u\}) > 0$.

3b Let S be a random subset of vertices obtained by including each vertex with probability $1/2$ independently of other vertices. Show that

$$\mathbb{E}[f(S)] = |E|/4 \geq \mathrm{OPT}/4\,,$$

where $\mathrm{OPT} = \max_{T\subseteq V} f(T)$.

Also give an example of a graph where $|E| = \mathrm{OPT}$ and thus it shows that the analysis is tight with respect to OPT.

Solution: Let O be an optimal solution and let $E' = \{(u,v) \in E : u \in O, v \notin O\}$. Then $\mathrm{OPT} = f(O) = |E'|$. For each $(u,v) \in E'$, let $X_{u,v}$ be the random variable such that, $X_{u,v} = 1$ if $u \in S$ and $v \notin S$, and $X_{u,v} = 0$ otherwise. Then we have that $f(S) \geq \sum_{(u,v)\in E'} X_{u,v}$. By the choice of S, we have that $\mathbb{E}[X_{u,v}] = \Pr[X_{u,v} = 1] = \Pr[u \in S] \cdot \Pr[v \notin S] = \frac{1}{2} \cdot \frac{1}{2} = 1/4$ for all $(u,v) \in E'$. Therefore, by linearity of expectation, we have that

$$\mathbb{E}[f(S)] \geq \sum_{(u,v)\in E'} \mathbb{E}[X_{u,v}] = \frac{1}{4}|E'| = \frac{\mathrm{OPT}'}{4}.$$

To see this is tight, consider the graph $G = (V = [2n], E = \{1,3,\ldots,2n-1\} \times \{2,4,\ldots,2n\})$. It has $2n$ vertices and all arcs goes from an odd vertex to an even vertex so that $OPT = |E|$ and the set $O = \{1,3,\ldots,2n-1\}$ of all odd vertices is an optimal set. In this case, we can easily verify that $f(S) = \sum_{(u,v)\in E} X_{u,v}$ and $\mathbb{E}[f(S)] = |E|/4 = \mathrm{OPT}/4$.

3c (*) Consider any submodular function f that is

- non-negative: $f(T) \geq 0$ for all T.

Let S be a random subset of elements obtained by including each element with probability $1/2$ independently of other elements. Then

$$\mathbb{E}[f(S)] \geq \mathrm{OPT}/4\,,$$

where $\mathrm{OPT} = \max_T f(T)$.

This shows that the simple randomized algorithm actually gives a good approximation to any (even non-monotone) submodular function assuming it is non-negative.

Solution: Let O be an optimal set so that $\mathrm{OPT} = f(O)$. Let $O(1/2)$ be a random subset of O (i.e., each element $o \in O$ is selected independently with probability $1/2$). Recall from Problem 4 that $\mathbb{E}[f(O(1/2))] \geq (1/2)f(O) + (1-1/2)f(\emptyset) \geq (1/2)\mathrm{OPT}$. Let $\bar{O} = N \setminus O$ and let $\bar{O}(1/2)$ be a random subset of $\bar{O}$. Let $O' \subset O$. The function $g_{O'}(T) = f(T \cup O')$ is also a submodular function. By the same argument, we have that $\mathbb{E}[g_{O'}(\bar{O}(1/2))] \geq (1-1/2)g_{O'}(\bar{O}) + (1/2)g_{O'}(\emptyset) \geq (1/2)f(\emptyset \cup O') = (1/2)f(O')$. Thus if $N(1/2)$ is a random subset of N, we have the following:

$$\begin{aligned}
\mathbb{E}[f(N(1/2))] &= \sum_{O'\subseteq O} \Pr[O(1/2) = O']\mathbb{E}[f(N(1/2)) \,|\, O(1/2) = O'] \\
&= \sum_{O'\subseteq O} \Pr[O(1/2) = O']\mathbb{E}[g_{O'}(\bar{O}(1/2))] \\
&\geq \frac{1}{2}\sum_{O'\subseteq O} \Pr[O(1/2) = O']f(O') \\
&= \frac{1}{2}\mathbb{E}[f(O(1/2))] \geq \frac{1}{4}\mathrm{OPT}.
\end{aligned}$$