Homework 1

Submission: June 13 in class

1. Basic Properties of Submodular Functions:
   Let \( f, g : 2^\mathcal{N} \to \mathbb{R}_+ \) be submodular functions over ground set \( \mathcal{N} \). Prove the following basic properties of submodular functions:
   
   (a) If \( h(S) \triangleq f(S) + g(S) \) for every \( S \subseteq \mathcal{N} \), then \( h \) is submodular.
   
   (b) For every fixed \( T \subseteq \mathcal{N} \), if \( h(S) \triangleq f(S \cap T) \) for every \( S \subseteq \mathcal{N} \) then \( h \) is submodular.
   
   (c) If \( h(S) \triangleq f(\mathcal{N} \setminus S) \) for every \( S \subseteq \mathcal{N} \), then \( h \) is submodular.
   
   (d) If \( f \) is monotone, then for every \( c \geq 0 \) the function defined by \( h(S) \triangleq \min\{c, f(S)\} \) for every \( S \subseteq \mathcal{N} \) is monotone and submodular.

2. Faster Greedy by Sampling
   Let \( f : 2^\mathcal{N} \to \mathbb{R}_+ \) be a monotone submodular functions over ground set \( \mathcal{N} \), and let \( k \in \mathbb{N} \) be a cardinality bound. We are interested in the following problem: \( \max \{ f(S) : S \subseteq \mathcal{N}, |S| \leq k \} \). Consider the following sampling based algorithm for the above problem parameterized by \( \varepsilon \):
   
   - \( S_0 \leftarrow \emptyset \).
   - For \( i = 1 \) to \( k \) do:
     - Let \( M_i \subseteq \mathcal{N} \) be a uniformly random subset of size \( \left\lceil \frac{n \ln(1/\varepsilon)}{k} \right\rceil \).
     - Let \( u_i \in M_i \) be the element maximizing the marginal value \( f(S_{i-1} \cup \{u_i\}) - f(S_{i-1}) \).
     - \( S_i \leftarrow S_{i-1} \).
   - Return \( S_k \).

   (a) Show that the running time of the algorithm is \( O(n \ln(1/\varepsilon)) \).

   (b) Let \( S^* \) be some optimal solution to the problem. The goal is to prove that for every iteration \( i = 1, \ldots, k \):
      \[
      \mathbb{E}[f(S_i) - f(S_{i-1})] \geq \frac{1 - \varepsilon}{k} (f(S^*) - \mathbb{E}[f(S_{i-1})]) \quad (*)
      \]

   Fix \( i \) and let \( v_1, \ldots, v_k \) be the \( k \) elements sorted according to the marginal values:
      \[
      f(S_{i-1} \cup \{v_1\}) - f(S_{i-1}) \geq \ldots \geq f(S_{i-1} \cup \{v_k\}) - f(S_{i-1}) .
      \]

   Let \( X_j \) be the indicator for the event that \( M_i \cap \{v_1, \ldots, v_j\} \neq \emptyset \).
      i. Prove that:
      \[
      f(S_{i-1} \cup \{u_i\}) - f(S_{i-1}) = \sum_{j=1}^{k-1} X_j \left( (f(S_{i-1} \cup \{v_j\}) - f(S_{i-1})) - (f(S_{i-1} \cup \{v_{j+1}\}) - f(S_{i-1})) \right) + X_k (f(S_{i-1} \cup \{v_k\}) - f(S_{i-1}))
      \]
ii. Prove that: $E[X_j] \geq 1 - \left(1 - \frac{\ln(1/\varepsilon)}{k}\right)^j$ for every $1 \leq j \leq k$.

iii. Prove $(\ast)$.\(^1\)

(c) Assuming $\varepsilon \leq 1 - 1/e$ solve $(\ast)$ and prove that $E[f(S_k)] \geq (1 - 1/e - \varepsilon) f(S^*)$.

\(^1\)Hint: use Chebyshev's sum inequality, for $a_1 \geq \ldots \geq a_n$ and $b_1 \geq \ldots \geq b_n$ the following holds:

$$\frac{1}{n} \sum_{\ell=1}^{n} a_\ell b_\ell \geq \left( \frac{1}{n} \sum_{\ell=1}^{n} a_\ell \right) \left( \frac{1}{n} \sum_{\ell=1}^{n} b_\ell \right)$$