Submodular Optimization (236621) Spring 2022/2023

July 3, 2023

1. Prove or refute that the following set functions $f : 2^N \to \mathbb{R}_+$ are submodular.

   (a) For every $i = 1,\ldots,k$ let $g_i : 2^N \to \mathbb{R}_+$ be k submodular functions over N, and let $a_1,\ldots,a_k$ be k non-negative coefficients. For every $S \subseteq N$ define $f(S) = a_1 g_1(S) + \ldots + a_k g_k(S)$.

   (b) Let $G = (V,E)$ be a directed graph, $N = V$, and $g(S)$ is the number of directed edges $(u \to v) \in E$ such that $u \in S$ and $v \notin S$. Additionally, let $b \in \mathbb{N}$ be a given threshold. For every $S \subseteq V$ define $f(S) = \min\{g(S), b\}$.

   (c) Let $g : 2^U \to \mathbb{R}_+$ be s submodular function over U. Choose the ground set N as follows $N = \{u : u \in U\} \cup \{u' : u \in U\}$, *i.e.*, N is the disjoint union of two independent copies of U. For every $S \subseteq N$ denote by $S_U \subseteq U$ the following set $\{u \in U : u \in S \text{ or } u' \in S\}$. Define $f(S) = g(S_U)$ for every $S \subseteq N$.

   (d) Let $g : 2^N \to \mathbb{R}_+$ be a submodular function over N and let $T \subseteq N$. Define $f(S) = g((T \setminus S) \cup (S \setminus T))$ for every $S \subseteq N$, *i.e.*, $f(S)$ equals g on the reflection of S with respect to T (all elements of S that are in T are swapped and elements of S outside of T are not swapped).

   (e) Let $G = (V,E)$ be an undirected graph equipped with non-negative edge weights $b : E \to \mathbb{R}_+$. For every vertex $u \in V$ denote by $B(u)$ the total weight of edges touching u, *i.e.*, $B(u) = \sum_{e=(u,v)\in E} b_e$. Let $\{\theta_u\}_{u\in V}$ be independent identically distributed uniform random variables in $[0,1]$. Given an initial $A_0 \subseteq V$ of infected vertices, consider the following dynamics which progresses in discrete steps: in the i^{th} step let A_i be,

   $$A_i = A_{i-1} \cup \left\{ u \notin A_{i-1} : \sum_{e=(u,v)\in E : v\in A_{i-1}} b_e \geq \theta_u B(u) \right\}.$$

   Denote by i_{A_0} the step after which no new infected vertices are added, *i.e.*, $A_{i_{A_0}} = A_{i_{A_0}+1} = A_{i_{A_0}+2} = \ldots$ (note that i_{A_0} is a random variable whose distribution depends on the initial infected set A_0). Set the ground set N to be V and for every $S \subseteq V$ define $f(S) = \mathbb{E}\left[|A_{i_S}|\right]$.

   Hint: prove that f is submodular.

2. Let $f : 2^N \to \mathbb{R}_+$ be a monotone submodular functions over a ground set N, and let $k \in \mathbb{N}$ be a cardinality bound. We are interested in solving: $\max\{f(S) : S \subseteq N, |S| \leq k\}$. The goal of this question is to obtain an algorithm that for every $\varepsilon > 0$ achieves an (almost) tight approximation of $1 - 1/e - \varepsilon$ in running time of $O(n \ln(1/\varepsilon))$. Consider the following sampling based algorithm parameterized by $\varepsilon > 0$:

   - $S_0 \leftarrow \emptyset$.
   - For $i = 1$ to k do:

- Let $M_i \subseteq N$ be a uniformly random subset of size $\lceil \frac{n \ln(1/\varepsilon)}{k} \rceil$.
- $u_i \leftarrow \text{argmax}\{f(S_{i-1} \cup \{u\}) - f(S_{i-1}) : u \in M_i\}$.
- $S_i \leftarrow S_{i-1} \cup \{u_i\}$.

- 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 current goal is to prove that for every iteration $i = 1, \dots, k$ the following holds:

$$\mathbb{E}[f(S_i) - f(S_{i-1})] \geq \frac{1-\varepsilon}{k}(f(S^*) - \mathbb{E}[f(S_{i-1})]). \qquad (*)$$

Fix i and let $v_1, \dots, v_k$ be the k elements of S^* 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 $M_i \cap \{v_1, \dots, v_j\} \neq \emptyset$. Follow the following steps to prove $(*)$.

i. Prove that the following inequality always holds (note that S_{i-1}, X_j and u_i are random variables):

$$f(S_{i-1} \cup \{u_i\}) - f(S_{i-1}) \geq \sum_{j=1}^{k-1} X_j \left[\left(f(S_{i-1} \cup \{v_j\}) - f(S_{i-1}) \right) - \left(f(S_{i-1} \cup \{v_{j+1}\}) - f(S_{i-1}) \right) \right] + X_k \left(f(S_{i-1} \cup \{v_k\}) - f(S_{i-1}) \right)$$

ii. Prove that: $\mathbb{E}[X_j] \geq 1 - \left(1 - \frac{\ln(1/\varepsilon)}{k}\right)^j$ for every $j = 1, \dots, k$.

iii. Prove $(*)$.

Hint: use Chebyshev's sum inequality, which states that for $a_1 \geq \ldots \geq a_n$ and $b_1 \geq \ldots \geq b_n$ the following holds:

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

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

3. Let $f : 2^N \to \mathbb{R}_+$ be a submodular function, which is not necessarily monotone, over ground set N. Let $M = (N, I)$ be a partition matroid of rank k where $N = N_1 \cup N_2 \cup \ldots \cup N_k$ is the partition of the ground set N that is associated with the matroid. We are interested in solving $\max\{f(S) : S \subseteq N, S \in I\}$. The goal of this question is obtaining a fast combinatorial algorithm that achieves an approximation of $1/4$ for the problem. Consider the following algorithm that is parameterized by $t \in \mathbb{N}$, which indicates how many iterations it performs:

- $S_0 \leftarrow \emptyset$ and $J_0 \leftarrow \emptyset$.
- for $i = 1$ to t do:
  - $\forall j = 1, \dots, k$: $u_j^i \leftarrow \text{argmax}\{f(S_{i-1} \cup \{u\}) - f(S_{i-1}) : u \in N_j\}$.
  - Let j_i be a uniformly random number from $\{1, \dots, k\}$.
  - $S_i \leftarrow S_{i-1}$ and $J_i \leftarrow J_{i-1}$.

  - if $j_i \notin J_{i-1}$ then:
    * $J_i \leftarrow J_i \cup \{j_i\}$.
    * if $f(S_{i-1} \cup \{u^i_{j_i}\}) - f(S_{i-1}) \geq 0$ then: $S_i \leftarrow S_i \cup \{u^i_{j_i}\}$.
- return S_t.

(a) Show that $S_i \in I$ for every $i = 1, \ldots, t$.

(b) For every $S \in I$ denote by O_S the best feasible completion of S, *i.e.*,

$$O_S = \text{argmax}\{f(S \cup T) : T \subseteq N \setminus S, S \cup T \in I\}.$$

Prove that for every $i = 1, \ldots, t$:

$$\mathbb{E}[f(S_i)] \geq \left(1 - \frac{1}{k}\right) \mathbb{E}[f(S_{i-1})] + \frac{1}{k} \mathbb{E}\left[f(S_{i-1} \cup O_{S_{i-1}})\right].$$

(c) Prove that for every $i - 1, \ldots, t$:

$$\mathbb{E}[f(S_i \cup O_{S_i})] \geq \left(1 - \frac{2}{k}\right) \mathbb{E}\left[f(S_{i-1} \cup O_{S_{i-1}})\right].$$

(d) Solve the above recursive formula and prove that for every $i = 1, \ldots, t$:

$$\mathbb{E}[f(S_i)] \geq \left(\left(1 - \frac{1}{k}\right)^i - \left(1 - \frac{2}{k}\right)^i\right) f(S^*).$$

Here S^* is some optimal solution to the problem.

Hint: note that without loss of generality $O_\emptyset = S^*$.

(e) Show that there is a choice of $t \in \mathbb{N}$ such that $\mathbb{E}[f(S_t)] \geq 1/4 \cdot f(S^*)$.

4. Prove the following properties of continuous extensions of submodular functions.

(a) Recall that the concave closure $f^+ : [0,1]^N \to \mathbb{R}_+$ of a set function $f : 2^N \to \mathbb{R}_+$ is defined as follows:

$$f^+(\mathbf{x}) \triangleq \max\{\sum_{S \subseteq N} \alpha_S f(S) : \sum_{S \subseteq N} \alpha_S = 1, \alpha_S \geq 0 \; \forall S \subseteq N, \sum_{S \subseteq N} \alpha_S \mathbf{1}_S = \mathbf{x}\} \quad \forall \mathbf{x} \in [0,1]^N$$

Prove that there exists $\mathbf{x} \in [0,1]^N$ for which it is NP-hard to compute $f^+(\mathbf{x})$ assuming f is submodular.

Hint: reduce from the Max-Cut problem.

(b) Let $F : [0,1]^N \to \mathbb{R}_+$ be the multilinear extensions of a submodular set function $f : 2^N \to \mathbb{R}_+$. Prove that for every $\mathbf{x} \in [0,1]^N$ and non-negative direction vector $\mathbf{d} \in \mathbb{R}^N_+$, *i.e.*, $d_i \geq 0$ for every $i \in N$, the function $g(s) \triangleq F(\mathbf{x} + s\mathbf{d})$ is concave in s (here $s \in \mathbb{R}$).