Maximization of Submodular Set Functions

Biomedical Signal Processing, Imaging, Reasoning, and Learning (BSPIRAL) Group

Author: Jamshid Sourati

Reviewers: Murat Akacayya
Yeganeh M. Marghi
Paula Gonzalez

Supervisors: Jennifer G. Dy
Deniz Erdogmus

January 2015

* Cite this report in the following format:
Jamshid Soruati, Maximization of Submodular Set Functions, Technical Report, BSPIRAL-150111-R0, Northeastern University, 2015.
Abstract

In this technical report, we aim to give a simple yet detailed analysis of several various submodular maximization algorithms. We start from analyzing the classical greedy algorithm, firstly discussed by Nemhauser et al. (1978), that guarantees a tight bound for constrained maximization of monotonically submodular set functions. We then continue by discussing two randomized algorithms proposed recently by Buchbinder et. al, for constrained and non-constrained maximization of nonnegative submodular functions that are not necessarily monotone.
Contents

1 Introduction 3

2 Preliminaries 4

3 Maximization Algorithms 8
   3.1 Monotone Functions ................................. 8
   3.2 Non-Monotone Functions ............................ 10
       3.2.1 Unconstrained Maximization .................... 11
       3.2.2 Constrained Maximization ..................... 16
1 Introduction

Combinatorial optimization plays an important role in many real applications of different fields such as machine learning. It frequently happens that the encountered combinatorial optimization is an NP-hard problem, hence there is no efficient algorithm to find its global solution. Relaxation to continuous space is a popular approach to escape from NP-hardness, however, it requires employing heuristics for discretizing back the obtained continuous solution. Fortunately, in many cases, the combinatorial objective turns out to be submodular (Krause et al., 2008; Jegelka and Bilmes, 2011; Kim et al., 2011), a family of set functions that shares both convexity and concavity properties of continuous functions (Lovász, 1983). Nevertheless, they do not indicate similar behavior for different optimization problems, specifically they are easy to minimize, but NP-hard to maximize (Bach, 2013). There is a yet growing literature on designing efficient algorithms for the latter problem, that is optimizing for a local maximum of such functions.

Throughout this paper we assume that $\mathcal{U}$ is a finite set and values of the set function $f$ are provided by an oracle for any subsets of $\mathcal{U}$:

**Definition 1.** A set function $f : 2^\mathcal{U} \to \mathbb{R}$ is said to be submodular iff

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B), \quad \forall A, B \subseteq \mathcal{U}$$

(1)

Monotonicity of submodular functions is an important property based on which one can dichotomize the maximization algorithms.

**Definition 2.** The set function $f : 2^\mathcal{U} \to \mathbb{R}$ is said to be monotone (nondecreasing) iff for every $S \subseteq T \subseteq \mathcal{U}$ we have $f(S) \leq f(T)$.

Here we consider both constrained and unconstrained maximization problems. Note that there are different types of constraints in this context, but we focus only on those that bound the cardinality of the solution from above.

**Problem 1. (Unconstrained)** Let $f : 2^\mathcal{U} \to \mathbb{R}$ be a submodular set function. Find the subset maximizing $f$: $\arg \max_{A \subseteq \mathcal{U}} f(A)$.

**Problem 2. (Constrained)** Let $f : 2^\mathcal{U} \to \mathbb{R}$ be a submodular set function. Find the subset with a given maximum size $k > 0$ maximizing $f$: $\arg \max_{A \subseteq \mathcal{U}, |A| \leq k} f(A)$.

As solving these cases are NP-hard, we always have to approximate the solution as close as possible to the real maximum. Every approximating algorithm is to be analyzed to show that the (expected) objective value at the output is no less than a fraction of the true maximum value of the set function. Clearly, this fraction is between 0 and 1, and it is desired to be as large as possible.

**Definition 3.** Suppose $f : 2^\mathcal{U} \to \mathbb{R}$ is a set function with its maximum occurred at $\mathcal{A}^* \subseteq \mathcal{U}$. An algorithm is said to be $\alpha$-approximation of the maximization problem (where $0 \leq \alpha \leq 1$ is called the approximation factor), iff its output $\mathcal{A}$ is guaranteed to satisfy $f(\mathcal{A}) \geq \alpha f(\mathcal{A}^*)$. 


For randomized algorithms, the definition above would be the same over the expected value of the output, i.e. having $\mathbb{E}[f(A)] \geq \alpha f(A^*)$.

Almost all of the proposed maximization algorithms are either deterministic or randomized iterative approaches that try to approximately locate the maximum of the objective. Existing algorithms can be categorized into the following groups, based on the properties of the objective function and the optimization constraints:

1. **Constrained maximization of monotone submodular set functions**— Nemhauser et al. (1978) proposed a $(1 - \frac{1}{e})$-approximation greedy algorithm for maximizing monotone submodular functions with cardinality constraints. This algorithm is proven to capture the optimal approximation factor in this case, assuming that values of $f$ are provided by an oracle for any subsets of $\mathcal{U}$. This optimality is in the sense that if a better approximation is to be achieved, the set function needs to be evaluated at exponentially many subsets (Nemhauser and Wolsey, 1978).

2. **Maximization of non-negative submodular set functions**— Maximizing non-monotone functions is harder and most of the work made the pre-assumption of non-negativity for the set functions.

   (a) **Without Constraints**— Feige et al. (2011) made one of the first attempts to design efficient algorithms and prove the corresponding lower-bounds for unconstrained maximization of non-negative functions that are not necessarily monotone. In addition to proposing a $(2/5)$-approximation algorithm he also proved that the optimal approximation factor in this case is $1/2$. The hard results (optimal approximation factors) in both monotone and non-monotone cases are re-derived using a joint framework by Vondrák (2013) in which the so-called *multilinear continuous extension* of the set functions is used. More recently Buchbinder et al. (2012) gave a $(1/3)$-approximation deterministic and $(1/2)$-approximation randomized algorithms for unconstrained maximization of nonnegative submodular functions.

   (b) **With cardinality constraints**— Moreover, cardinality constrained maximization is also recently considered for nonnegative functions that are not necessarily monotone by Buchbinder et al. (2014) in form of a random discrete and a deterministic continuous algorithms.

In this technical report, we provide the main theoretical results related to each item listed above. Starting by the most basic algorithm proposed by Nemhauser et al. (1978) for maximizing monotone submodular set functions, we then discuss the randomized algorithm proposed by Buchbinder et al. (2012) for unconstrained maximizing nonnegative submodular set functions. Finally we give the randomized discrete algorithm from the work by Buchbinder et al. (2014) for constrained maximization of nonnegative submodular functions.

### 2 Preliminaries

Before going through details of the algorithms, here, we give some general lemmas and propositions that will be useful in the later analysis. Lemma 1 shows subadditivity of submodular
functions; Proposition 1, together with Definition 4, establishes an equivalent definition of
submodularity; and Lemmas 2 and 3 prove two other useful properties of submodular functions.

Lemma 1. Let \( f : 2^\mathcal{U} \to \mathbb{R} \) be a nonnegative submodular function and \( A \subseteq \mathcal{U} \) a nonempty
subset with member-wise representation \( A = \{a_1, \ldots, a_m\} \). Then \( f(A) \leq \sum_{i=1}^{m} f(\{a_i\}) \).

Proof. By applying submodularity inequality to \( f(A) \) iteratively we get:

\[
\begin{align*}
f(\{a_1, \ldots, a_m\}) & \leq f(\{a_1, \ldots, a_{m-1}\}) + f(\{a_m\}) - f(\emptyset) \\
& \leq f(\{a_1, \ldots, a_{m-2}\}) + f(\{a_{m-1}\}) + f(\{a_m\}) - 2f(\emptyset) \\
& \vdots \\
& \leq \sum_{i=1}^{m} f(\{a_i\}) - (m-1)f(\emptyset).
\end{align*}
\]

Nonnegativity of \( f \) implies that \( f(\emptyset) \geq 0 \) and therefore \( f(\{a_1, \ldots, a_m\}) \leq \sum_{i=1}^{m} f(\{a_i\}) \). \( \square \)

Generally speaking, definition (1) is not easy to use immediately. It is more common to
work with an equivalent definition based on the discrete derivative function \( \rho_f \):

Definition 4. Let \( f : 2^\mathcal{U} \to \mathbb{R} \) be a set function, \( A \subseteq \mathcal{U} \) and \( u \in \mathcal{U} \). The discrete derivative
of function \( f \) is defined as

\[
\rho_f(A,u) = f(A \cup \{u\}) - f(A). \tag{2}
\]

Proposition 1. Let \( f : 2^\mathcal{U} \to \mathbb{R} \) be a set function. \( f \) is submodular iff we have

\[
\rho_f(S,u) \geq \rho_f(T,u), \quad \forall S \subseteq T \subseteq \mathcal{U}, u \in \mathcal{U} - T \tag{3}
\]

Proof. \((\Rightarrow)\): In definition 1 take \( A = S \cup \{u\} \) and \( B = T \):

\[
\begin{align*}
f(S \cup \{u\}) + f(T) & \geq f(S \cup \{u\} \cup T) + f((S \cup \{u\}) \cap T) \\
& = f(T \cup \{u\}) + f(S) \\
\Rightarrow \quad f(S \cup \{u\}) - f(S) & \geq f(T \cup \{u\}) - f(T) \\
\Rightarrow \quad \rho_f(S,u) & \geq \rho_f(T,u).
\end{align*}
\]

\((\Leftarrow)\): Take \( A, B \subseteq \mathcal{U} \). If \( A \subseteq B \) we get \( f(A \cup B) + f(A \cap B) = f(A) + f(B) \) and
the submodularity condition shown in (1) holds. Now assume \( A \nsubseteq B \) hence \( A - B \neq \emptyset \).
As \( \mathcal{U} \) is finite \( A - B \subseteq \mathcal{U} \) has a finite number of members and can be represented by
\( A - B = \{u_1, \ldots, u_{|A - B|}\} \). Now define the sets \( T_i \) and \( S_i \) for \( i = 0, \ldots, |A - B| - 1 \) as the following

\[
\begin{align*}
T_i &= \begin{cases} 
B, & i = 0 \\
B \cup \{u_1, \ldots, u_i\}, & i = 1, \ldots, |A - B| - 1
\end{cases} \\
S_i &= \begin{cases} 
A \cap B, & i = 0 \\
(A \cap B) \cup \{u_1, \ldots, u_i\}, & i = 1, \ldots, |A - B| - 1
\end{cases}
\end{align*}
\]
Notice that $S_{i-1} \subseteq T_{i-1}$ and $u_i \notin T_{i-1} \Rightarrow u_i \in \mathcal{U} - T_{i-1}$ for any $i = 1, \ldots, |A - B|$ hence equation (3) can be applied:

$$\rho_f(S_{i-1}, u_i) \geq \rho_f(T_{i-1}, u_i), \quad i = 1, \ldots, |A - B|$$

Summing both sides of the inequalities above yields:

$$\sum_{i=1}^{|A - B|} \rho_f(S_{i-1}, u_i) \geq \sum_{i=1}^{|A - B|} \rho_f(T_{i-1}, u_i)$$

$$\Rightarrow \sum_{i=1}^{|A - B|} f(S_{i-1} \cup \{u_i\}) - f(S_{i-1}) \geq \sum_{i=1}^{|A - B|} f(T_{i-1} \cup \{u_i\}) - f(T_{i-1})$$

$$\Rightarrow f(S_{|A - B|-1} \cup \{u_{|A - B|}\}) - f(S_0) \geq f(T_{|A - B|-1} \cup u_{|A - B|}) - f(T_0),$$

where we used the equality $S_{i-1} \cup \{u_i\} = S_i$ and the telescopic properties of the summation of the inequalities. Hence:

$$f((A \cap B) \cup \{u_1, \ldots, u_{|A - B|}\}) - f(A \cap B) \geq f(B \cup \{u_1, \ldots, u_{|A - B|}\}) - f(B)$$

$$\Rightarrow f((A \cap B) \cup (A - B)) - f(f \cap B) \geq f(B \cup (A - B)) - f(B)$$

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

\[ \square \]

We will also require the following lemmas in our later derivations:

**Lemma 2.** If $f : 2^\mathcal{U} \to \mathbb{R}$ is a submodular set function and $B \subseteq \mathcal{U}$, then the set function $g_B : 2^\mathcal{U} \to \mathbb{R}$, defined as $g_B(A) = f(A \cup B), \forall A \subseteq \mathcal{U}$ is also submodular for any $B \subseteq \mathcal{U}$.

**Proof.** Take a subset $B \subseteq \mathcal{U}$. It suffices to show that $\rho_{g_B}(S, u) \geq \rho_{g_B}(T, u), \forall S \subseteq T \subseteq \mathcal{U}$ and $u \notin T$. We consider the following two cases:

(I) $u \in B$:

$$\rho_{g_B}(S, u) = g_B(S \cup \{u\}) - g_B(S)$$

$$= f(S \cup B \cup \{u\}) - f(S \cup B)$$

$$= f(S \cup B) - f(S \cup B) = 0.$$

Similarly we can verify that $\rho_{g_B}(T, u) = g_B(T \cup \{u\}) - g_B(T) = 0$ and thus the equality over the discrete derivatives holds. Therefore, $\rho_{g_B}(S, u) \geq \rho_{g_B}(T, u)$ is satisfied.

(II) $u \notin B$:

$$\rho_{g_B}(S, u) - \rho_{g_B}(T, u) = [g_B(S \cup \{u\}) - g_B(S)] - [g_B(T \cup \{u\}) - g_B(T)]$$

$$= [f(S \cup B \cup \{u\}) - f(S \cup B)] - [f(T \cup B \cup \{u\}) - f(T \cup B)]$$

$$= \rho_f(S \cup B, u) - \rho_f(T \cup B, u) \geq 0,$$

where the last inequality follows directly from submodularity of $f$. 

6
Lemma 3. If $f : 2^U \to \mathbb{R}$ is a submodular set function, then for every $A, B \subseteq U$ we have

$$f(A) \leq f(B) + \sum_{u \in A - B} \rho_f(B, u) - \sum_{u \in B - A} \rho_f(A \cup B - \{u\}, u)$$

(4)

Proof. Take two arbitrary sets $A, B \subseteq U$. First assume $A \nsubseteq B$ hence $A - B \neq \emptyset$. As $A - B$ is finite represent it by $A - B = \{w_1, ..., w_{|A - B|}\}$ and write $f(A \cup B) - f(B)$ in terms of its elements using the following telescopic sum:

$$f(A \cup B) - f(B) = \sum_{i=1}^{|A - B|} \left[ f \left( B \cup \left\{ \bigcup_{j=1}^i w_j \right\} \right) - f \left( B \cup \left\{ \bigcup_{j=1}^{i-1} w_j \right\} \right) \right]$$

$$= \sum_{i=1}^{|A - B|} \rho_f \left( B \cup \left\{ \bigcup_{j=1}^{i-1} w_j \right\}, w_i \right)$$

$$\leq \sum_{i=1}^{|A - B|} \rho_f(B, w_i) = \sum_{w \in A - B} \rho_f(B, w),$$

(5)

where the last inequality is obtained by applying inequality (3) to all discrete derivatives inside the summation with $T = B \cup \left\{ \bigcup_{j=1}^{i-1} w_j \right\}$, $S = B \subset T$ and $y = w_i \notin B = T \Rightarrow y \in U - T$ (for all $i = 1, ..., |A - B|$). Also note that inequality (5) is valid even if $A \subseteq B$, since $f(A \cup B) - f(B) = f(B) - f(B) = 0 = \sum_{w \in A - B = \emptyset} \rho_f(B, w)$.

Now assume $B \nsubseteq A$ hence $B - A \neq \emptyset$. Represent it by $B - A = \{z_1, ..., z_{|B - A|}\}$ and similar to the previous case use telescopic sums properties:

$$f(A \cup B) - f(A) = \sum_{i=1}^{|B - A|} \left[ f \left( A \cup \left\{ \bigcup_{j=1}^i z_j \right\} \right) - f \left( A \cup \left\{ \bigcup_{j=1}^{i-1} z_j \right\} \right) \right]$$

$$= \sum_{i=1}^{|B - A|} \rho_f \left( A \cup \left\{ \bigcup_{j=1}^{i-1} z_j \right\}, z_i \right)$$

$$\geq \sum_{i=1}^{|B - A|} \rho_f(A \cup B - \{z_i\}, z_i) = \sum_{z \in B - A} \rho_f(A \cup B - \{z\}, z),$$

(6)

where the inequality here is derived by applying inequality (3) with $T = A \cup B - \{z_i\}$, $S = A \cup \{z_j\}_{j=1}^{i-1} \subseteq T$ and $y = z_i \notin A \cup B - \{z_i\} = T$ (for all $i = 1, ..., |B - A|$). The yielded inequality in (6) holds even when $B \subseteq A$ since $f(A \cup B) - f(A) = f(A) - f(A) = 0 = \sum_{z \in B - A = \emptyset} \rho_f(A \cup B - \{z\}, z)$.

Finally multiply $-1$ to both sides of inequality (6) and add them to both sides of inequality (5) to get:

$$f(A) - f(B) \leq \sum_{w \in A - B} \rho_f(B, w) - \sum_{z \in B - A} \rho_f(A \cup B - \{z\}, z)$$

$$\Rightarrow f(A) \leq f(B) + \sum_{w \in A - B} \rho_f(B, w) - \sum_{z \in B - A} \rho_f(A \cup B - \{z\}, z).$$
Algorithm 1: Greedy approach for solving Problem 2 for monotone and submodular functions

**Inputs:** the source set $U$, the submodular set function $f$, the maximum size of the subset $k$

**Outputs:** a subset $A \subseteq U$ of size at most $k$

1. $A_0 \leftarrow \emptyset$
2. $U_0 \leftarrow U$
3. for $t = 1 \rightarrow k$ do
   4. $u_t \leftarrow \arg \max_{u \in U_{t-1}} f(A_{t-1} \cup \{u\})$
   5. $A_t \leftarrow A_{t-1} \cup \{u_t\}$
   6. $U_t \leftarrow U - A_t$
4. return $A = A_k$

We are now ready to start analyzing some of the algorithms for submodular maximizing.

## 3 Maximization Algorithms

In the remaining of this report, we analyze in detail some of the existing maximization algorithms that belong to the categories listed above.

### 3.1 Monotone Functions

In this section, we focus on the size-constrained maximization of monotone and submodular set functions (look at the categorization given in section 1), that is solving Problem 2 when the function $f$ satisfies inequality (1) and the monotonicity condition in Definition 2. Here we discuss the efficient iterative algorithm that is firstly introduced in the seminal paper by Nemhauser et al. (1978).

The pseudocode of the approach is shown in Algorithm 1. The proof of this well-known result presented here, is based on the previous works (Nemhauser et al., 1978; Krause and Golovin, 2012).

Suppose $A$ denotes the output of the algorithm and $A_t$ the set generated after $t$ iterations ($0 \leq t \leq k$) and $A^*$ the optimal solution, i.e. $A^* = \arg \max_{A \subseteq U, |A| \leq k} f(A)$.

**Theorem 1.** Let $f : 2^U \rightarrow \mathbb{R}$ be a monotone (nondecreasing) submodular set function, $A^*$ be the optimal solution and $A \subseteq U$ be the subset obtained through Algorithm 1. If $A_t$ is the
intermediate result of this algorithm at iteration $t$, then

$$f(A^*) \leq f(A_t) + k \cdot \rho_f(A_t, u_{t+1}), \quad 0 \leq t < k,$$

(7)

Proof. Use Lemma 3 with $A = A^*$ and $B = A_t$ to write:

$$f(A^*) \leq f(A_t) + \sum_{w \in A^* - A_t} \rho_f(A_t, w) - \sum_{z \in A_t - A^*} \rho_f(A^* \cup A_t - \{z\}, z)$$

where the last term is obtained by noticing that $f$ is nondecreasing hence

$$\rho_f(A^* \cup A_t - \{z\}, z) \geq 0 \Rightarrow -\rho_f(A^* \cup A_t - \{z\}, z) \leq 0, \forall z \in U.$$  

Also note that at each iteration $0 \leq t < k$, $u_{t+1}$ is selected by maximizing $f(A_t \cup \{u\})$, which is equivalent to maximizing $\rho_f(A_t, u)$ with respect to $u$. Therefore, one can write $\rho_f(A_t, w) \leq \rho_f(A_t, u_{t+1}), \forall w \in U_t = U - A_t$. Applying this to the inequality above together with the fact that $|A^* - A_t| \leq |A^*| \leq k$ complete the proof:

$$f(A^*) \leq f(A_t) + \sum_{w \in A^* - A_t} \rho_f(A_t, w)$$

$$\leq f(A_t) + k \cdot \rho_f(A_t, u_{t+1}).$$

(8)

Theorem 1 implies that if at any iteration $0 \leq t \leq k - 1$ we have $f(A_t) = f(A_{t+1})$, i.e. $\rho_f(A_t, u_{t+1}) = 0$, the output of Algorithm 1 is equal to the optimal solution:

**Corollary 1.** Let $f : 2^U \rightarrow \mathbb{R}$ be a monotone (nondecreasing) submodular set function, $A^*$ be the optimal solution and $A \subseteq U$ be the subset obtained through Algorithm 1. If there exists an iteration $0 \leq t < k$ such that $\rho_f(A_t, u_{t+1}) = 0$, then $f(A_t) = f(A^*)$.

Proof. From Theorem 1 we have:

$$f(A^*) \leq f(A_t) + k \cdot \rho_f(A_t, u_{t+1}) = f(A_t).$$

(9)

On the other hand, $A^*$ is assumed to be the optimal solution, suggesting $f(A^*) \geq f(A_t)$, hence $f(A_t) = f(A^*)$.  

Now we prove the main result providing a lower bound for the value of $f$ at the greedily obtained solution in terms of the optimal value:

**Theorem 2.** Let $f : 2^U \rightarrow \mathbb{R}$ be a submodular and monotone (nondecreasing) set function with $f(\emptyset) = 0$. If $A$ is the output of Algorithm 1 and $A^*$ is the optimal solution to Problem 2, then we have:

$$f(A) \geq \left[1 - \left(\frac{k-1}{k}\right)^k\right] f(A^*) \geq \left(1 - \frac{1}{e}\right) f(A^*).$$

(10)
Proof. From Theorem 1 we have (for 0 ≤ t < k):

\[ f(\mathcal{A}^*) - f(\mathcal{A}_t) \leq k \cdot \rho_f(\mathcal{A}_t, u_{t+1}) \]
\[ = k \cdot [f(\mathcal{A}_t \cup \{u_{t+1}\}) - f(\mathcal{A}_t)] \]
\[ = k \cdot [f(\mathcal{A}_{t+1}) - f(\mathcal{A}_t)] \]
\[ = k \cdot [f(\mathcal{A}^*) - f(\mathcal{A}_t)] - k \cdot [f(\mathcal{A}^*) - f(\mathcal{A}_{t+1})] \]

Taking \( f(\mathcal{A}^*) - f(\mathcal{A}_t) \) terms to one side and simplifying the expression yields:
\[ \frac{f(\mathcal{A}^*) - f(\mathcal{A}_{t+1})}{f(\mathcal{A}^*) - f(\mathcal{A}_t)} \leq \frac{k - 1}{k}. \]

Now let us relate \( f(\mathcal{A}^*), f(\mathcal{A}) \) and \( f(\mathcal{A}_0 = \emptyset) \):

\[ \frac{f(\mathcal{A}^*) - f(\mathcal{A})}{f(\mathcal{A}^*) - f(\mathcal{A}_0)} = \prod_{i=0}^{\lvert \mathcal{A} \rvert-1} \frac{f(\mathcal{A}^*) - f(\mathcal{A}_{i+1})}{f(\mathcal{A}^*) - f(\mathcal{A}_i)} \leq \left( \frac{k - 1}{k} \right)^{\lvert \mathcal{A} \rvert} \leq \left( 1 - \frac{1}{k} \right)^k, \]

from which we can write:

\[ f(\mathcal{A}) \geq \left[ 1 - \left( 1 - \frac{1}{k} \right)^k \right] f(\mathcal{A}^*) + \left( 1 - \frac{1}{k} \right)^k f(\emptyset) \]
\[ = \left[ 1 - \left( 1 - \frac{1}{k} \right)^k \right] f(\mathcal{A}^*), \]

where we used the assumption that \( f(\emptyset) = 0 \). Finally we use the inequality \( 1 - x \leq e^{-x}, \forall x \in \mathbb{R}^+ \) to get:

\[ \left[ 1 - \left( 1 - \frac{1}{k} \right)^k \right] f(\mathcal{A}^*) \geq \left( 1 - e^{-\frac{k}{k}} \right) f(\mathcal{A}^*) \]

yielding

\[ f(\mathcal{A}) \geq \left( 1 - \frac{1}{e} \right) f(\mathcal{A}^*). \]

\[ \square \]

3.2 Non-Monotone Functions

In this section, we discuss algorithms for solving Problems 1 and 2 for submodular and non-negative set functions.

Maximization of non-monotone submodular functions is not as straightforward as the monotone ones. There are recent studies on greedy approaches similar to Algorithm 1 for the family of non-negative and submodular functions which are not necessarily monotone, but the lower bounds that are derived for the proposed algorithms are not as good as cases where monotonicity is present. For instance, as we saw in the last section, if \( f \) is monotone and submodular, its constrained maximizer \( f(\mathcal{A}^*) \) (i.e. the solution to Problem 2) can be approximated with the approximation factor of \( \left( 1 - \frac{1}{e} \right) \). However, it is shown that when there
is no constraints, no efficient algorithm can maximize non-negative, symmetric \(^1\) and submodular set functions (which is possibly non-monotone) with an approximation factor better than \(\frac{1}{2}\) (Feige et al., 2011). Moreover, Algorithm 1 suggests that the bound \((1 - \frac{1}{e})\) could be attained by the submodular and monotone maximization in a deterministic framework. But, we will see shortly that the proposed algorithms analyzed here for maximizing submodular and non-monotone functions, achieve their relatively weaker factors by introducing randomness, and therefore their bound achievements are in the sense of expectation.

Here we discuss two randomized greedy algorithms for solving unconstrained and constrained optimization problems, proposed by Buchbinder et al. (2012) and Buchbinder et al. (2014). We will see that these algorithms, shown in Algorithms 2 and 3, result constant \((1/2)\) and \((1/e)\) approximation factors for Problems 1 and 2 respectively. In addition we also analyze a constrained version of Algorithm 2.

### 3.2.1 Unconstrained Maximization

Pseudocode of the greedy algorithm proposed for unconstrained maximization is shown in Algorithm 2 (Buchbinder et al., 2012). It is a randomized double algorithm in the sense that it starts with two solutions \(A_0 = \emptyset\) and \(B_0 = U\) and at each iteration with some probability adds an element to the former or removes it from the latter. These probabilities depend on how much improvement will be attained by taking either of the actions. At the end of such procedure, the two evolving sets \(A_t\) and \(B_t\) would coincide for \(t = n\) and either of them can be returned as the ultimate solution.

For obtaining a lower-bound for the objective value of the output of Algorithm 2 we need the following two lemmas:

**Lemma 4.** Let \(f : 2^U \to \mathbb{R}\) be a submodular function and \(A_{t-1}\) and \(B_{t-1}\) the intermediate double random sets generated up to the \(t\)-th iteration of Algorithm 2. If \(a_t\) and \(b_t\) are defined as in lines 4 and 5 of Algorithm 2, then we have \(a_t + b_t \geq 0\) for every iteration \(1 \leq t \leq n\).

**Proof.** Denote \(U\) member-wise by \(U = \{u_1, ..., u_n\}\). By the way we initialize and then continue sampling the double random sets, at each iteration of the algorithm we have \(B_{t-1} = A_{t-1} \cup \{u_t, ..., u_n\}\) hence \(A_{t-1} \subseteq B_{t-1}\). We consider the two sets \(A_{t-1} \cup \{u_t\}\) and \(B_{t-1} - \{u_t\}\) that are involved in \(a_t\) and \(b_t\) and see that their union and intersection are \(B_{t-1}\) and \(A_{t-1}\) respectively:

\[
(A_{t-1} \cup \{u_t\}) \cup (B_{t-1} - \{u_t\}) = A_{t-1} \cup (\{u_t\} \cap B_{t-1})
= A_{t-1} \cup B_{t-1} = B_{t-1}
\]

\[
(A_{t-1} \cup \{u_t\}) \cap (B_{t-1} - \{u_t\}) = A_{t-1} \cap (B_{t-1} - \{u_t\})
= A_{t-1} \cap (A_{t-1} \cup \{u_{t+1}, ..., u_n\})
= A_{t-1}.
\]

\(^1\)The set function \(f : 2^U \to \mathbb{R}\) is called symmetric if

\[
f(A) = f(U - A), \quad \forall A \subseteq U.
\]
Algorithm 2: Randomized double greedy approach for solving problem 1 for non-negative submodular functions

**Inputs:** the source set \( \mathcal{U} = \{u_1, ..., u_n\} \), the submodular set function \( f \)

**Outputs:** a subset \( \mathcal{A} \subseteq \mathcal{U} \)

```plaintext
/* Initializations */
1 \( \mathcal{A}_0 \leftarrow \emptyset \)
2 \( \mathcal{B}_0 \leftarrow \mathcal{U} \)

/* Starting the Iterations */
for \( t = 1 \rightarrow n \) do
4 \( a_t \leftarrow f(\mathcal{A}_{t-1} \cup \{u_t\}) - f(\mathcal{A}_{t-1}) \)
5 \( b_t \leftarrow f(\mathcal{B}_{t-1} - \{u_t\}) - f(\mathcal{B}_{t-1}) \)
6 \( a'_t \leftarrow \max(a_t, 0) \)
7 \( b'_t \leftarrow \max(b_t, 0) \)

/* Randomized Updating of the Double Loop Variables */
with probability \( a'_t / (a'_t + b'_t) \) do
9 \( \mathcal{A}_t \leftarrow \mathcal{A}_{t-1} \cup \{u_t\} \)
10 \( \mathcal{B}_t \leftarrow \mathcal{B}_{t-1} \)
with probability \( b'_t / (a'_t + b'_t) \) do
12 \( \mathcal{A}_t \leftarrow \mathcal{A}_{t-1} \)
13 \( \mathcal{B}_t \leftarrow \mathcal{B}_{t-1} - \{u_t\} \)

return \( \mathcal{A} = \mathcal{A}_n = \mathcal{B}_n \)
```

* If \( a'_t + b'_t = 0 \), take \( a'_t / (a'_t + b'_t) = 1. \)

Forming \( a_t + b_t \) and using definition of submodularity (equation (1)) results:

\[
    a_t + b_t = f(\mathcal{A}_{t-1} \cup \{u_t\}) - f(\mathcal{A}_{t-1}) + f(\mathcal{B}_{t-1} - \{u_t\}) - f(\mathcal{B}_{t-1}) \geq 0.
\]

Now define the set \( \mathcal{C}_t = (\mathcal{A}^* \cup \mathcal{A}_t) \cap \mathcal{B}_t \) where \( \mathcal{A}^* \) is the optimal solution of the unconstrained maximization. Note that \( \mathcal{C}_0 = \mathcal{A}^* \), the optimal set, and \( \mathcal{C}_n = \mathcal{A}_n = \mathcal{B}_n \), the output. The set sequence \( \mathcal{C}_0 = \mathcal{A}^*, \mathcal{C}_1, ..., \mathcal{C}_n = \mathcal{A}_n \) generated by each iteration of the algorithm, starts from the optimal set and ends up with the output of the algorithm. Similarly we can look at the sequence of objective values \( f(\mathcal{A}^*), f(\mathcal{C}_1), ..., f(\mathcal{C}_n) \). The following analysis is done by bounding the expected amount of decrease in the objective when moving from the global maximum \( f(\mathcal{C}_0 = \mathcal{A}^*) \) towards the output value \( f(\mathcal{C}_n) \) in the sequence \( \{f(\mathcal{C}_i)\}_{i=0}^n \).

First we put a bound over the expected decrease of objective between two consecutive iterations. Note that in what follows for analysis of algorithm 2, all the expectations are jointly taken with respect to the random sets \( \mathcal{A}_t \) and \( \mathcal{B}_t \):

**Lemma 5.** Let \( f : 2^\mathcal{U} \rightarrow \mathbb{R} \) be a submodular function and \( \mathcal{A}_{t-1} \) and \( \mathcal{B}_{t-1} \) be the intermediate double random sets generated up to the t-th iteration of Algorithm 2 (1 \( \leq t \leq n \)). If we
define \( C_t := (A^* \cup A_t) \cap B_t \) where \( A^* \) is the optimal solution to Problem 1, then:

\[
\mathbb{E}[f(C_{t-1}) - f(C_t)] \leq \frac{1}{2} \mathbb{E}[f(A_t) - f(A_{t-1}) + f(B_t) - f(B_{t-1})].
\]  

(12)

\textbf{Proof.} We consider all possible cases in every iteration \( 1 \leq t \leq n \) and show that the inequality (12) holds in all of them. These cases are separated based on the signs of \( a_t \) and \( b_t \) and also whether the next element \( u_t \) is inside \( A^* \) or not. From Lemma 4, \( a_t + b_t \geq 0 \) hence they cannot be negative at the same time and we get the following cases (as mentioned before \( B_{t-1} = A_{t-1} \cup \{u_t, \ldots, u_n\} \)):

(I) \( (a_t \geq 0, b_t \leq 0) \):

This case implies that with probability one \( A_t = A_{t-1} \cup \{u_t\} \) and \( B_t = B_{t-1} \) hence the right-hand-side of the inequality is \( \frac{1}{2}[f(A_t) - f(A_{t-1})] = a_t/2. \) We also get

\[
C_t = (A^* \cup A_t) \cap B_t = (A^* \cup A_{t-1} \cup \{u_t\}) \cap B_{t-1} = C_{t-1} \cup (\{u_t\} \cap B_{t-1}) = C_{t-1} \cup \{u_t\}.
\]

The left-hand-side of the inequality depends on the following two cases:

(i) \( (u_t \in A^*) \):  

In this case \( u_t \in C_{t-1} \) implying that \( C_t = C_{t-1} \) and the left-hand-side becomes zero. Hence the inequality holds.

(ii) \( (u_t \notin A^*) \)

The right-hand-side in this case is \( f(C_{t-1}) - f(C_{t-1} \cup \{u_t\}) = -\rho_f(C_{t-1}, u_t). \) By submodularity of \( f \) we get \( \rho_f(C_{t-1}, u_t) \geq \rho_f(B_{t-1} - \{u_t\}, u_t) \) because \( C_{t-1} \subseteq B_{t-1} - \{u_t\} \) and \( u_t \notin B_{t-1} - \{u_t\} \). This implies that

\[
f(C_{t-1}) - f(C_t) = -\rho_f(C_{t-1}, u_t) \leq -\rho_f(B_{t-1} - \{u_t\}, u_t) = f(B_{t-1} - \{u_t\}) - f(B_{t-1}) = b_t \leq 0.
\]

Recall that we saw the right-hand-side is \( \frac{a_t}{2} \geq 0 \) hence the inequality holds.

(II) \( (a_t < 0, b_t \geq 0) \):

This case is similar to the previous one. Here with probability one we have \( A_t = A_{t-1} \) and \( B_t = B_{t-1} - \{u_t\} \) implying that the right-hand-side becomes \( \frac{1}{2}[f(B_t) - f(B_{t-1})] = b_t/2. \) We also get:

\[
C_t = (A^* \cup A_t) \cap B_t = (A^* \cup A_{t-1}) \cap (B_{t-1} - \{u_t\}) = C_{t-1} - \{u_t\}.
\]

For obtaining the left-hand-side we consider the two following cases:

(i) \( (u_t \in A^*) \):
Having \( u_t \) inside \( \mathcal{A}^* \) suggests that \( u_t \in \mathcal{C}_{t-1} \) (because it exists in both \( \mathcal{A}^* \) and \( \mathcal{B}_{t-1} \)). The left-hand-side can be written as \( \rho_f(\mathcal{C}_{t-1} - \{u_t\}, u_t) \). Note that since \( \mathcal{A}_{t-1} \subseteq \mathcal{B}_{t-1} \) we get:

\[
\mathcal{C}_{t-1} = (\mathcal{A}^* \cup \mathcal{A}_{t-1}) \cap \mathcal{B}_{t-1} \\
= (\mathcal{A}^* \cap \mathcal{B}_{t-1}) \cup (\mathcal{A}_{t-1} \cap \mathcal{B}_{t-1}) \\
= (\mathcal{A}^* \cap \mathcal{B}_{t-1}) \cup \mathcal{A}_{t-1} \supseteq \mathcal{A}_{t-1}
\]

\[ \Rightarrow \mathcal{C}_{t-1} - \{u_t\} \supseteq \mathcal{A}_{t-1} - \{u_t\} = \mathcal{A}_{t-1}, \]

since \( u_t \notin \mathcal{A}_{t-1} \). Now by submodularity of \( f \) we have

\[
\rho_f(\mathcal{C}_{t-1} - \{u_t\}, u_t) \leq \rho_f(\mathcal{A}_{t-1}, u_t) \\
= f(\mathcal{A}_{t-1} \cup \{u_t\}) - f(\mathcal{A}_{t-1}) \\
= a_t < 0.
\]

But recall the right-hand-side was equal to \( b_t/2 \geq 0 \) hence the inequality holds.

(ii) \( (u_t \notin \mathcal{A}^*) \):

In this case \( u_t \) is not in \( \mathcal{A}^* \) nor in \( \mathcal{A}_{t-1} \) implying that \( u_t \notin \mathcal{C}_{t-1} \), therefore the left-hand-side is zero and clearly less than or equal to \( b_t/2 \).

(III) \( (a_t \geq 0, b_t > 0) \):

In this case both probabilities are non-zero and we have a non-degenerate discrete distribution over the possible actions. Note that here \( a'_t = a_t \) and \( b'_t = b_t \) thus the probabilities are \( \frac{a_t}{a_t + b_t} \) and \( \frac{b_t}{a_t + b_t} \) resulting the following:\(^2\)

\[
\mathbb{E}[f(\mathcal{A}_t) - f(\mathcal{A}_{t-1}) + f(\mathcal{B}_t) - f(\mathcal{B}_{t-1})] = \]

\[
= \left( \frac{a_t}{a_t + b_t} \right) \left[ f(\mathcal{A}_{t-1} \cup \{u_t\}) - f(\mathcal{A}_{t-1}) \right] + \\
\left( \frac{b_t}{a_t + b_t} \right) \left[ f(\mathcal{B}_{t-1} - \{u_t\}) - f(\mathcal{B}_{t-1}) \right] \\
= \frac{a_t^2 + b_t^2}{a_t + b_t}.
\]

where the first term corresponds to adding \( u_t \) to \( \mathcal{A}_{t-1} \) and the second term corresponds to removing it from \( \mathcal{B}_{t-1} \). We already mentioned that \( \mathcal{C}_t \) would be equal to \( \mathcal{C}_{t-1} \cup \{u_t\} \) and \( \mathcal{C}_{t-1} - \{u_t\} \) in each case, respectively, and:

\[
\mathbb{E}[f(\mathcal{C}_{t-1}) - f(\mathcal{C}_t)] = \left( \frac{a_t}{a_t + b_t} \right) \left[ f(\mathcal{C}_{t-1}) - f(\mathcal{C}_{t-1} \cup \{u_t\}) \right] + \\
\left( \frac{b_t}{a_t + b_t} \right) \left[ f(\mathcal{C}_{t-1}) - f(\mathcal{C}_{t-1} - \{u_t\}) \right]
\]

Now note that the two terms shown above cannot be nonzero simultaneously and one of them is going to be zero depending on whether \( u_t \) is inside \( \mathcal{A}^* \) (hence inside \( \mathcal{C}_{t-1} \)):

\(^2\)Note that in this case \( b_t \) is considered a strictly positive number. The case where both numbers might become zero is considered in case 1, since in if that happens we assume \( \frac{a_t}{a_t + b_t} = 1 \) as if \( a_t > 0 \) and \( b_t \leq 0 \).
(i) \((u_t \in \mathcal{A}^\ast)\):

\[
\mathbb{E}[f(C_{t-1}) - f(C_t)] = \left(\frac{b_t}{a_t + b_t}\right) [f(C_{t-1}) - f(C_{t-1} - \{u_t\})] \\
\leq \frac{a_t b_t}{a_t + b_t},
\]

where the inequality is already shown in case II(i).

(ii) \((u_t \notin \mathcal{A}^\ast)\):

\[
\mathbb{E}[f(C_{t-1}) - f(C_t)] = \left(\frac{a_t}{a_t + b_t}\right) [f(C_{t-1}) - f(C_{t-1} \cup \{u_t\})] \\
\leq \frac{a_t b_t}{a_t + b_t},
\]

where the inequality is already shown in case 1(b).

Therefore one can always claim that \(\mathbb{E}[f(C_{t-1}) - f(C_t)] \leq \frac{a_t b_t}{a_t + b_t}\). Finally notice that \(\frac{2a_t b_t}{a_t + b_t} \leq \frac{a_t^2 + b_t^2}{a_t + b_t}\) for any \(a_t\) and \(b_t\) to complete the proof:

\[
\mathbb{E}[f(C_{t-1}) - f(C_t)] \leq \frac{a_t b_t}{a_t + b_t} \leq \frac{1}{2} \left(\frac{a_t^2 + b_t^2}{a_t + b_t}\right) \\
= \frac{1}{2} \mathbb{E}[f(\mathcal{A}_t) - f(\mathcal{A}_{t-1}) + f(\mathcal{B}_t) - f(\mathcal{B}_{t-1})]
\]

\[\square\]

**Theorem 3.** Let \(f : 2^\mathcal{U} \rightarrow \mathbb{R}\) be a submodular set function. If \(\mathcal{A}_n\) is the output of Algorithm 2 and \(\mathcal{A}^\ast\) is the optimal solution of Problem 1, then

\[
\mathbb{E}[f(\mathcal{A}_n)] \geq \frac{1}{2} f(\mathcal{A}^\ast). \quad (13)
\]

**Proof.** The inequality (12) is proven for all iterations \(1 \leq t \leq n\), summing them all will result a telescopic sum in each side:

\[
\sum_{t=1}^{n} \mathbb{E}[f(C_{t-1}) - f(C_t)] \leq \frac{1}{2} \sum_{t=1}^{n} \mathbb{E}[f(\mathcal{A}_t) - f(\mathcal{A}_{t-1}) + f(\mathcal{B}_t) - f(\mathcal{B}_{t-1})] \\
\Rightarrow \mathbb{E}[f(C_0) - f(C_n)] \leq \frac{1}{2} \mathbb{E}[f(\mathcal{A}_n) + f(\mathcal{B}_n) - f(\mathcal{A}_0) - f(\mathcal{B}_0)].
\]

Now remember \(C_0 = \mathcal{A}^\ast, A_0 = \emptyset, B_0 = \mathcal{U}\) and \(C_n = \mathcal{A}_n = \mathcal{B}_n\):

\[
f(\mathcal{A}^\ast) \leq 2\mathbb{E}[f(\mathcal{A}_n)] - f(\emptyset) - f(\mathcal{U}) \\
\leq 2\mathbb{E}[f(\mathcal{A}_n)].
\]

\[\square\]
3.2.2 Constrained Maximization

In this section, first we give discuss a slight deviation of Algorithm 2, which makes it constrained, and then analyze a separate single-variable randomized algorithm recently proposed for constrained maximization of nonnegative and submodular set functions.

In the constrained version of algorithm 2, we terminate the algorithm as soon as the size of the set $A_t$ reaches the maximum size $k > 0$ determined in the constrained problem. Therefore, if this happens at iteration $t^*$ the output would be $A_{t^*}$. Clearly, $t^*$ can be at least $k$ (if in all iterations we add elements into $A_t$) and at most $n$ (if $|A_n| \leq k$). We have the following result on the lower bound of $\mathbb{E}[f(A_{t^*})]$ where we used the implicit assumption of upper-boundedness of $f$ over $U$:

**Theorem 4.** Let $f : 2^U \rightarrow \mathbb{R}$ be a submodular function, $A_{t^*}$ be the output of the constrained version of Algorithm 2 and $A^*$ be the optimal solution of Problem 1. If we denote the upper bound of $f$ over $U$ by $M$, i.e. $f\{u\} \leq M \ \forall u \in U$, then

$$\mathbb{E}[f(A_{t^*})] \geq \frac{1}{2} f(A^*) - (n-k)M. \ (14)$$

**Proof.** First we find an upper bound for the difference between the expected value of $f(A_n)$ and $A_{t^*}$. We know that $A_{t^*} \subseteq A_n$. Denote the set difference by $S_{t^*} = A_n - A_{t^*}$. Note that if $S_{t^*} = \emptyset$ then we have $t^* = n$ (since $A_{t^*} = A_n$) and therefore inequality (14) holds by Theorem 3. Thus assume that $S_{t^*} \neq \emptyset$ and denote it member-wise by $S_{t^*} = \{s_1, ..., s_m\}$. Clearly, $S_{t^*} \subseteq \{u_{t^*+1}, ..., u_n\}$ and therefore $m \leq n - t^* \leq n - k$. From submodularity of $f$ and also using Lemma 1 we have:

$$f(A_n) - f(A_{t^*}) = f(A_{t^*} \cup S_{t^*}) - f(A_{t^*})$$

$$\leq f(A_{t^*}) + f(S_{t^*}) - f(\emptyset) - f(A_{t^*})$$

$$\leq f(S_{t^*}) \leq \sum_{i=1}^{m} f(\{s_i\})$$

$$\leq mM \leq (n-k)M$$

This inequality together with Theorem 3 suggests that

$$\mathbb{E}[f(A_{t^*})] \geq \mathbb{E}[f(A_n)] - (n-k)M$$

$$\geq \frac{1}{2} f(A^*) - (n-k)M.$$ 

Now we turn to the recent algorithm proposed by Buchbinder et al. (2014) for maximizing nonnegative submodular functions with cardinality constraint. The pseudocode shown in Algorithm 3 is very similar to Algorithm 1 with a certain degree of added exploration, i.e. instead of selecting the element with highest $f$ at each iteration, a sample is chosen randomly from the set of $k$ elements with the highest $f$ values. Construction of $M_t$ in line 4 can be done simply by sorting the elements in $U_{t-1}$ in a descending order in terms of their $f$
Algorithm 3: Randomized greedy approach for solving Problem 2 for non-negative and submodular functions

**Inputs:** the source set $U$, the submodular set function $f$, the maximum size of the subset $k$

**Outputs:** a subset $A \subseteq U$ of size at most $k$

```plaintext
/* Initializations */
1 $A_0 \leftarrow \emptyset$
2 $U_0 \leftarrow U$

/* Starting the Iterations */
3 for $t = 1 \rightarrow k$
4     $M_t \leftarrow \arg\max_{u \in M \subseteq U \atop |M|=k} \sum_{u \in M} \rho_f(A_{t-1}, u)$
5     $u_t \leftarrow \text{RANDOM}(M_t)$
6     $A_t \leftarrow A_{t-1} \cup \{u_t\}$
7     $U_t \leftarrow U - A_t$
8 return $A = A_k$
```

values, and picking the first $k$ elements. Line 5 randomly selects one of the entries from the constructed $M_t$.

Extra care should be given to the size of $U_t$ in this algorithm. Note that if $|U| < 2k$, after $k$ iterations we would run out of enough samples to put inside $M_t$. The following Remark discuss this issue:

**Remark 1.** In algorithm 3, we can always assume that $|U| \geq 2k$.

*Proof.* If $k > \frac{|U|}{2}$ we can define a dummy set $D$ with $2k$ elements such that for every $A \subseteq U$ we have $f(A) = f(A - D)$. By doing this we might get some of the elements of $D$ in the output of the algorithm, we simply remove them after the algorithm is finished.

Note that introducing this set does not effect our following proofs and the output of the algorithm since it does not change neither $f(A^*)$ nor $f(A_t), 0 \leq t \leq k$. $\square$

Observe that according to the algorithm size of the output might be less than $k$ only if $n < 2k$ which introduces some dummy elements to the universe set.

Buchbinder et al. (2014) discuss that the performance of Algorithm 3 is expected to be similar to Algorithm 1 when the function to maximize is monotone. That is, while it guarantees the weaker bound of $\frac{1}{e}$ for non-monotone functions, it maintains the stronger bound of $\left(1 - \frac{1}{e}\right)$ if its input set function is monotone. Of course, in contrary to Algorithm 1, because of the randomness here, this bound is achieved in expectation:

- $f$ monotone and submodular: $\mathbb{E}_A[f(A)] \geq \left(1 - \frac{1}{e}\right)f(A^*)$ (15a)
- $f$ non-negative and submodular: $\mathbb{E}_A[f(A)] \geq \frac{1}{e}f(A^*)$ (15b)
In this section, first we show (15a) in Theorem 5, and then build up the proof for (15b).

**Theorem 5.** Let \( f : 2^U \to \mathbb{R} \) be a submodular and monotone set function with \( f(\emptyset) = 0 \). If \( \mathcal{A} \) is the output of Algorithm 3 and \( \mathcal{A}^* \) is the optimal solution of Problem 2, then

\[
\mathbb{E}_\mathcal{A}[f(\mathcal{A})] \geq \left( 1 - \frac{1}{e} \right) \cdot f(\mathcal{A}^*). \tag{16}
\]

**Proof.** Suppose \( \mathcal{A}_{t-1} \) hence \( \mathcal{M}_t \) is fixed (for some \( 1 \leq t \leq k \)). In this case \( u_t \) is a discrete random variable distributed uniformly over its domain \( \mathcal{M}_t \). Also observe that \( \mathcal{M}_t \) gives the set of \( k \) elements in \( U_{t-1} \) which has the highest discrete derivatives, that is \( \sum_{u \in \mathcal{M}_t} \rho f(\mathcal{A}_{t-1}, u) \geq \sum_{u \in S} \rho f(\mathcal{A}_{t-1}, u), \forall S \subseteq U_{t-1}, |S| \leq k \). Thus one can write:

\[
\mathbb{E}_{u_t \mid \mathcal{A}_{t-1}}[\rho f(\mathcal{A}_{t-1}, u_t)] = \frac{1}{k} \sum_{u_t \in \mathcal{M}_t} \rho f(\mathcal{A}_{t-1}, u_t) \geq \frac{1}{k} \sum_{u_t \in \mathcal{A}^* - \mathcal{A}_{t-1}} \rho f(\mathcal{A}_{t-1}, u_t). \tag{17}
\]

We know that \( \mathcal{A}^* - \mathcal{A}_{t-1} \subseteq U \) is finite, represent it by \( \mathcal{A}^* - \mathcal{A}_{t-1} = \{w_1, ..., w_r\}, r \geq 0 \). Submodularity of \( f \) implies that the following inequalities hold:

\[
\begin{align*}
\rho f(\mathcal{A}_{t-1}, w_1) & \geq \rho f(\mathcal{A}_{t-1}, w_1) & \text{(18a)} \\
\rho f(\mathcal{A}_{t-1}, w_2) & \geq \rho f(\mathcal{A}_{t-1} \cup \{w_1\}, w_2) & \text{(18b)} \\
\rho f(\mathcal{A}_{t-1}, w_3) & \geq \rho f(\mathcal{A}_{t-1} \cup \{w_1, w_2\}, w_3) & \text{(18c)} \\
\vdots \\
\rho f(\mathcal{A}_{t-1}, w_r) & \geq \rho f(\mathcal{A}_{t-1} \cup \{w_1, ..., w_{r-1}\}, w_r). & \text{(18d)}
\end{align*}
\]

Summing both sides of the inequalities in (18) yields:

\[
\sum_{u \in \mathcal{A}^* - \mathcal{A}_{t-1}} \rho f(\mathcal{A}_{t-1}, u) \geq f(\mathcal{A}_{t-1} \cup (\mathcal{A}^* - \mathcal{A}_{t-1})) - f(\mathcal{A}_{t-1})
\]

\[
= f(\mathcal{A}^* \cup \mathcal{A}_{t-1}) - f(\mathcal{A}_{t-1}) \tag{19}
\]

We substitute obtained inequality in (19) into (17) to get:

\[
\mathbb{E}_{u_t \mid \mathcal{A}_{t-1}}[\rho f(\mathcal{A}_{t-1}, u_t)] \geq \frac{1}{k} \cdot [f(\mathcal{A}^* \cup \mathcal{A}_{t-1}) - f(\mathcal{A}_{t-1})]
\]

\[
\geq \frac{1}{k} \cdot [f(\mathcal{A}^*) - f(\mathcal{A}_{t-1})], \tag{20}
\]

where the last inequality is obtained from the monotonicity of \( f \).

So far we had the assumption that the set \( \mathcal{A}_{t-1} \) was fixed. Unfix it and take the expectation of both sides of inequality (20) with respect to \( \mathcal{A}_{t-1} \). Then noticing that

\[
\mathbb{E}_{\mathcal{A}_{t-1}}[\mathbb{E}_{u_t \mid \mathcal{A}_{t-1}}[\rho f(\mathcal{A}_{t-1}, u_t)]] = \mathbb{E}_{\mathcal{A}_t}[\rho f(\mathcal{A}_{t-1}, u_t)] \]

\[
^3\text{For every two random variables } X \text{ and } Y \text{ and any joint function } f(X,Y) \text{ we have } \mathbb{E}_Y[\mathbb{E}_X[f(X,Y)]] = \mathbb{E}_{X,Y}[f(X,Y)].
\]
Lemma 6. Let $f : 2^U \to \mathbb{R}$ be a nonnegative submodular set function and $p : U \to [0, 1]$ be a probability assignment to the elements of $U$ with an upper bound $M$, i.e. $p(u) \leq M, \forall u \in U$. If we denote $S_p$ the set obtained by sampling from $U$ where the element $u \in U$ is selected with probability $p(u)$, that is $\Pr(u \in S_p) = p(u) \ \forall u \in U$, then we have:

$$\mathbb{E}_p[f(S_p)] \geq (1 - M)f(\emptyset). \quad (21)$$
Proof. Without loss of generality assume that the elements in \( \mathcal{U} \) are sorted in terms of their probability of being selected, i.e. \( \mathcal{U} = \{ u_1, \ldots, u_n \} \) such that \( p(u_i) \geq p(u_j), \forall 1 \leq i < j \leq n \). Also denote \( \mathcal{S}_p = \{ s_1, \ldots, s_{|\mathcal{S}_p|} \} \) and define \( \mathcal{U}_i := \{ u_1, \ldots, u_i \} \) for any \( 1 \leq i < n \) and \( \mathcal{U}_0 = \emptyset \). We can write \( f(\mathcal{S}_p) \) in terms of the telescopic sums over all the members of \( \mathcal{U} \):

\[
\begin{align*}
f(\mathcal{S}_p) &= f(\emptyset) + \sum_{i=1}^{|\mathcal{U}|} 1(u_i \in \mathcal{S}_p) \left[ f(\mathcal{U}_i \cap \mathcal{S}_p) - f(\mathcal{U}_{i-1} \cap \mathcal{S}_p) \right] \\
&= \sum_{i=1}^{|\mathcal{U}|} 1(u_i \in \mathcal{S}_p) \rho_f(\mathcal{U}_{i-1} \cap \mathcal{S}_p, u_i) \\
&\geq f(\emptyset) + \sum_{i=1}^{|\mathcal{U}|} 1(u_i \in \mathcal{S}_p) \rho_f(\mathcal{U}_{i-1}, u_i),
\end{align*}
\]

where the last inequality is true because of submodularity of \( f \). Now taking the expectation over both sides of the inequality above yields:

\[
\begin{align*}
\mathbb{E}_p[f(\mathcal{S}_p)] &\geq \mathbb{E}_p \left[ f(\emptyset) + \sum_{i=1}^{|\mathcal{U}|} 1(u_i \in \mathcal{S}_p) \rho_f(\mathcal{U}_{i-1}, u_i) \right] \\
&= f(\emptyset) + \sum_{i=1}^{|\mathcal{U}|} \mathbb{E}_p[1(u_i \in \mathcal{S}_p)] \rho_f(\mathcal{U}_{i-1}, u_i) \\
&= f(\emptyset) + \sum_{i=1}^{|\mathcal{U}|} p(u_i) \rho_f(\mathcal{U}_{i-1}, u_i),
\end{align*}
\]

where we used the fact that \( 1(u_i \in \mathcal{S}_p) \), the indicator function, is a random variable with Bernoulli distribution with parameter \( p(u_i) \), hence its expected value is \( p(u_i) \). Now we expand the discrete derivatives and use non-negativity of \( f \) to get:

\[
\begin{align*}
\mathbb{E}_p[f(\mathcal{S}_p)] &\geq f(\emptyset) + \sum_{i=1}^{|\mathcal{U}|} p(u_i)(f(\mathcal{U}_i) - f(\mathcal{U}_{i-1})) \\
&= [1 - p(u_1)]f(\emptyset) + \sum_{i=1}^{|\mathcal{U}|-1} p(u_i)f(\mathcal{U}_i) - \sum_{i=2}^{|\mathcal{U}|} p(u_i)f(\mathcal{U}_{i-1}) + p(u_{|\mathcal{U}|})f(\mathcal{U}) \\
&= [1 - p(u_1)]f(\emptyset) + \sum_{i=2}^{|\mathcal{U}|} [p(u_{i-1}) - p(u_i)]f(\mathcal{U}_{i-1}) + p(u_{|\mathcal{U}|})f(\mathcal{U}) \\
&\geq [1 - p(u_1)]f(\emptyset).
\end{align*}
\]

Finally recalling that \( M \) is an upper bound for the function \( p \), we can write

\[
\mathbb{E}_p[f(\mathcal{S}_p)] \geq [1 - p(u_1)]f(\emptyset) \geq (1 - M)f(\emptyset).
\]

\[\square\]
Lemma 7. Let \( f : 2^\mathcal{U} \to \mathbb{R} \) be a nonnegative submodular set function and \( \mathcal{A}_t \) and \( \mathcal{U}_t \) be the sets generated at the \( t \)-th iteration of Algorithm 3 for \( 0 \leq t \leq k \). If \( \mathcal{A}^* \) is the optimal solution of Problem 2, then we have:

\[
\mathbb{E}_{\mathcal{A}_t}[f(\mathcal{A}^* \cup \mathcal{A}_t)] \geq \left( 1 - \frac{1}{k} \right)^t f(\mathcal{A}^*). \tag{22}
\]

Proof. The claim is trivial for the case \( t = 0 \). Take \( 1 \leq t \leq k \) and \( u \notin \mathcal{A}_{t-1} \). Note \( \Pr(u_t = u|u \in \mathcal{M}_t) = \frac{1}{k} \) since \( u_t \) is selected randomly from \( \mathcal{M}_t \) which has size \( k \). The goal here is to compute \( \Pr(u \in \mathcal{A}_t) \) and then use the result of Lemma 6. For doing so, we start from computing the conditional probability that \( u \) will be selected in iteration \( t \) given that it is not selected so far. Note that the followings hold no matter if \( u \in \mathcal{A}^* \) or not:

\[
\Pr(u \in \mathcal{A}_t | u \notin \mathcal{A}_{t-1}) = \Pr(u \in \mathcal{M}_t, u = u_t | u \notin \mathcal{A}_{t-1}) = \Pr(u \in \mathcal{M}_t | u \notin \mathcal{A}_{t-1}) \cdot \Pr(u_t = u | u \in \mathcal{M}_t) \leq \frac{1}{k}.
\]

Hence:

\[
\Pr(u \notin \mathcal{A}_t | u \notin \mathcal{A}_{t-1}) \geq 1 - \frac{1}{k}.
\]

Now we compute the unconditional probability that \( u \in \mathcal{U} \) is inside \( \mathcal{A}_t \). However, it is easier to compute the probability of its complement, \( u \notin \mathcal{A}_t \):

\[
\Pr(u \notin \mathcal{A}_t) = \Pr(u \notin \mathcal{A}_1, \ldots, u \notin \mathcal{A}_t) = \Pr(u \notin \mathcal{A}_1) \cdot \Pr(u \notin \mathcal{A}_2 | u \notin \mathcal{A}_1) \cdots \Pr(u \notin \mathcal{A}_t | u \notin \mathcal{A}_{t-1}) \geq \left( 1 - \frac{1}{k} \right)^t,
\]

which gives us

\[
\Pr(u \in \mathcal{A}_t) \leq 1 - \left( 1 - \frac{1}{k} \right)^t.
\]

Now define the function \( g_{\mathcal{A}^*} : 2^\mathcal{U} \to \mathbb{R} \) as \( g_{\mathcal{A}^*}(A) = f(\mathcal{A}^* \cup A), \forall A \subseteq \mathcal{U} \). Note that we have \( g_{\mathcal{A}^*}(\mathcal{A}_t) = g_{\mathcal{A}^*}(\mathcal{A}_t - \mathcal{A}^*) = f(\mathcal{A}_t \cup \mathcal{A}^*) \) since \( f \) is a submodular set function, from Lemma 2, \( g_{\mathcal{A}^*} \) is also submodular, and we can use Lemma 6 with \( M = 1 - (1 - \frac{1}{k})^t \) to write:

\[
\mathbb{E}_{\mathcal{A}_t}[g_{\mathcal{A}^*}(\mathcal{A}_t)] = \mathbb{E}_{\mathcal{A}_t}[f(\mathcal{A}^* \cup \mathcal{A}_t)] \geq \left( 1 - \frac{1}{k} \right)^t g_{\mathcal{A}^*}(\emptyset) = \left( 1 - \frac{1}{k} \right)^t f(\mathcal{A}^*).
\]

Now we bring the main theorem which gives a lower bound on the expected value of the objective set function at the generated sets of Algorithm 3:

Theorem 6. Let \( f : 2^\mathcal{U} \to \mathbb{R} \) be a submodular nonnegative set function and \( \mathcal{A}_t \) and \( \mathcal{U}_t \) be the sets generated at the \( t \)-th iteration of Algorithm 3 for \( 0 \leq t \leq k \). If \( \mathcal{A}^* \) is the optimal solution of Problem 2, then we have:

\[
\mathbb{E}_{\mathcal{A}_t}[f(\mathcal{A}_t)] \geq \left( \frac{t}{k} \right) \cdot \left( 1 - \frac{1}{k} \right)^{t-1} f(\mathcal{A}^*), \quad 0 \leq t \leq k. \tag{23}
\]
Proof. The theorem is trivial for \( t = 0 \). Take \( 1 \leq t \leq k \) and fix \( A_{t-1} \) hence \( M_t \). In this case \( u_t \) is a discrete random variable distributed uniformly over its domain \( M_t \). Also observe that \( M_t \) gives the set of \( k \) elements in \( U_{t-1} \) which has the highest discrete derivatives, that is \( \sum_{u \in M_t} \rho_f(A_{t-1}, u) \geq \sum_{u \in S} \rho_f(A_{t-1}, u), \forall S \subseteq U_{t-1}, |S| \leq k \). Thus one can write:

\[
\mathbb{E}_{u_{t-1}|A_{t-1}}[\rho_f(A_{t-1}, u_{t-1})] = \frac{1}{k} \sum_{u_{t-1} \in M_t} \rho_f(A_{t-1}, u_{t-1}) \geq \frac{1}{k} \sum_{u_{t-1} \in A^* - A_{t-1}} \rho_f(A_{t-1}, u_{t-1}). \tag{24}
\]

We know that \( A^* - A_{t-1} \subseteq U \) is finite, represent it by \( A^* - A_{t-1} = \{w_1, ..., w_r\}, r \geq 0 \). Submodularity of \( f \) implies that the following inequalities hold:

\[
\begin{align*}
\rho_f(A_{t-1}, w_1) & \geq \rho_f(A_{t-1}, w_1) \\
\rho_f(A_{t-1}, w_2) & \geq \rho_f(A_{t-1} \cup \{w_1\}, w_2) \tag{25b} \\
\rho_f(A_{t-1}, w_3) & \geq \rho_f(A_{t-1} \cup \{w_1, w_2\}, w_3) \tag{25c} \\
& \vdots \tag{25d} \\
\rho_f(A_{t-1}, w_r) & \geq \rho_f(A_{t-1} \cup \{w_1, ..., w_{r-1}\}, w_r). \tag{25e}
\end{align*}
\]

Summing both sides of the inequalities in (25) and using the telescopic properties of the right-hand-side sum yields:

\[
\sum_{u \in A^* - A_{t-1}} \rho_f(A_{t-1}, u) \geq f(A_{t-1} \cup (A^* - A_{t-1})) - f(A_{t-1}) = f(A^* \cup A_{t-1}) - f(A_{t-1}) \tag{26}
\]

We substitute the inequality obtained in (26) into (24) to get:

\[
\mathbb{E}_{u_{t-1}|A_{t-1}}[\rho_f(A_{t-1}, u_{t-1})] \geq \frac{1}{k} \cdot [f(A^* \cup A_{t-1}) - f(A_{t-1})].
\]

So far we had the assumption that the set \( A_{t-1} \) was fixed. Unfix it and take the expectation of both sides with respect to \( A_{t-1} \). Then noticing that \( \mathbb{E}_{A_{t-1}}[\mathbb{E}_{u_{t-1}|A_{t-1}}[\rho_f(A_{t-1}, u_{t-1})]] = \mathbb{E}_{A_{t-1}}[\rho_f(A_{t-1}, u_{t-1})] \) we get:

\[
\mathbb{E}_{A_{t-1}}[\rho_f(A_{t-1}, u_{t-1})] \geq \frac{1}{k} \cdot \left[ \mathbb{E}_{A_{t-1}}[f(A^* \cup A_{t-1})] - \mathbb{E}_{A_{t-1}}[f(A_{t-1})] \right].
\]

Since \( f \) is a submodular and nonnegative set function we can use Lemma 7 to write:

\[
\mathbb{E}_{A_{t-1}}[\rho_f(A_{t-1}, u_{t-1})] \geq \frac{1}{k} \left[ \left( 1 - \frac{1}{k} \right)^{t-1} \cdot f(A^*) - \mathbb{E}_{A_{t-1}}[f(A_{t-1})] \right]. \tag{27}
\]

Having already mentioned that the proof is valid for \( t = 0 \), now we are ready to demonstrate an inductive proof for the theorem. Suppose (23) holds for \( 0 \leq t - 1 < k \), let us prove the inequality for \( t \):
First note that \( f(A_t) = f(A_{t-1}) + \rho f(A_{t-1}, u_t) \), take the expectation of both sides with respect to \( A_t \) and apply the inequality (27) and the inductive assumption \( \mathbb{E}_{A_{t-1}}[f(A_{t-1})] \geq \left(\frac{t-1}{k}\right) \cdot \left(1 - \frac{1}{k}\right)^{t-2} \cdot f(A^*) \) to get:

\[
\begin{align*}
\mathbb{E}_{A_t}[f(A_t)] & = \mathbb{E}_{A_{t-1}}[f(A_{t-1})] + \mathbb{E}_{A_t}[\rho f(A_{t-1}, u_t)] \\
& \geq \mathbb{E}_{A_{t-1}}[f(A_{t-1})] + \frac{1}{k} \left[ \left(1 - \frac{1}{k}\right)^{t-1} \cdot f(A^*) - \mathbb{E}_{A_{t-1}}[f(A_{t-1})] \right] \\
& = \left(1 - \frac{1}{k}\right) \mathbb{E}_{A_{t-1}}[f(A_{t-1})] + \frac{1}{k} \left(1 - \frac{1}{k}\right)^{t-1} \cdot f(A^*) \\
& \geq \left[ \left(1 - \frac{1}{k}\right) \cdot \left(1 - \frac{1}{k}\right)^{t-2} + \frac{1}{k} \left(1 - \frac{1}{k}\right)^{t-1} \right] \cdot f(A^*) \\
& = \left(\frac{t}{k}\right) \cdot \left(1 - \frac{1}{k}\right)^{t-1} \cdot f(A^*). \quad \square
\end{align*}
\]

**Corollary 2.** Let \( f : 2^U \to \mathbb{R} \) be a submodular nonnegative set function and \( A \) be the output of Algorithm 3. If \( A^* \) is the optimal solution Problem 2, then we have:

\[
\mathbb{E}_A[f(A)] \geq \frac{t}{e} f(A^*). \tag{28}
\]

**Proof.** \( f \) is a submodular and nonnegative set function hence we can use Theorem 6 with \( t = k \) to write:

\[
\mathbb{E}_{A_k}[f(A_k)] \geq \left(1 - \frac{1}{k}\right)^{k-1} \cdot f(A^*) \geq \frac{1}{e} f(A^*). \quad \square
\]

**References**


