Generalizing Performance Bounds for the Greedy Algorithm: Approximate Submodularity and Localization

Temitayo Ajayi *, Taewoo Lee †2 and Andrew Schaefer ‡1
1Rice University, Department of Computational and Applied Mathematics
2University of Houston, Department of Industrial Engineering

Abstract

The greedy algorithm is perhaps the simplest heuristic in combinatorial optimization. When maximizing a nonnegative, increasing, submodular function, the greedy algorithm has a worst-case performance bound proportional to its optimal objective value at a given cardinality. We introduce new metrics that quantify a function’s proximity to being submodular. Using our proposed criteria, we derive performance bounds for the greedy algorithm applied to non-submodular functions. Insights from our initial derivations allow us to generalize existing bounds, satisfy the criteria, and improve upon the numerical bounds. We examine multiple bounds using a generalization of the facility location problem that does not have a submodular objective function generally. In numerical examples, our new bounds are competitive with, and often an improvement on, those of Das and Kempe (2011), Horel and Singer (2016), and Zhou and Spanos (2016). We observe this collection of bounds provides complementary information about the optimized function.

1 Introduction

Although many combinatorial problems have polynomial-time algorithms, others are NP-hard and may lead to approximation algorithms that exploit special structure. The greedy algorithm is a natural heuristic for a wide variety of combinatorial problems (Nemhauser et al. [1978]). The greedy algorithm is highly intuitive, but its performance depends on the structure of the problem.

Submodularity of the optimized set function is a key property in many combinatorial optimization problems that allows for effective heuristics, including the greedy algorithm. Let Ω be a finite set of
elements. A function $f : 2^Ω \rightarrow \mathbb{R}$ is submodular if for any $\mathcal{A} \subseteq B \subseteq \Omega$ and $s \in \Omega \setminus B$, $f(B \cup \{s\}) - f(B) \leq f(\mathcal{A} \cup \{s\}) - f(\mathcal{A})$. Nemhauser et al. (1978) and Fisher et al. (1978) show that when maximizing a nonnegative, increasing, submodular function, the greedy algorithm is guaranteed to be within $1 - (1 - \frac{1}{L})^L$ of optimality, where $L$ is both the number of iterations run by the algorithm and the cardinality parameter. This has encouraged use of the greedy algorithm for combinatorial problems with submodular objectives.

The greedy algorithm is often used, with varying levels of success, on problems with no known theoretical guarantee. The greedy algorithm performs well empirically on graphical tagging (Christensen et al. 1995), protein structure recovery (Tuffery et al. 2005), and car sequencing (Gottlieb et al. 2003), although reasons for this are largely unexplored. Hence, the absence of performance guarantees does not itself prohibit successful heuristics.

Researchers have produced bounds for functions that are not necessarily increasing, nonnegative, and submodular. For example, Cornuéljols et al. (1977) derive a performance bound for the greedy algorithm on the uncapacitated facility location problem, whose objective function is submodular, but not necessarily nonnegative nor increasing. Krause et al. (2008) show that if a function is submodular and approximately monotonic the greedy algorithm has a performance bound. Still, their results require a submodular function.

Approximation guarantees for approximately submodular functions have attracted recent attention. Horel and Singer (2016) explore the trade-offs between additive and multiplicative bounds for the greedy algorithm’s performance on non-submodular functions. Horel and Singer (2016) note that if the range of the function is bounded, their “lower bounds and approximation results could equivalently be expressed for additive approximations of normalized functions.” Zhou and Spanos (2016) and Das and Kempe (2011) define metrics that they use to produce greedy algorithm performance bounds for approximately submodular functions. Unfortunately, some of these bounds are not well-defined in some simple cases (see Section 2.1).

In this paper, we study performance bounds for the greedy algorithm for optimization problems with non-submodular functions. To do so, we utilize metrics, including novel metrics and those from the literature, that quantify the worst-case violation of submodularity for a given function, which we then use to bound the performance of the greedy algorithm. Moreover, the bounds we present adhere to basic criteria that allow for flexibility in the search for solutions (e.g., sparse or dense). Our theoretical and numerical results provide insight into why the greedy algorithm can do well, even for problems with non-submodular objectives.

## 2 Greedy Algorithm Performance Guarantees

In this section, we present performance guarantees for the greedy algorithm on set function maximization subject to a cardinality constraint. Let $f : 2^\Omega \rightarrow \mathbb{R}_+$ be an increasing function, where $\Omega$ is a discrete,
finite set. Let \( \hat{S}_K \in \arg \max_{|S| \leq K} f(S) \), where \( K \) denotes the cardinality parameter, and \( \mathcal{S}_L \) as the set selected by the greedy algorithm at iteration \( L \) when maximizing the function. The remainder of this section reviews submodularity metrics and bounds in the literature and presents new bounds that offer different information. Omitted proofs of new results (other than corollaries) can be found in the appendix.

2.1 Existing Submodularity Metrics and Bounds

Existing performance bounds for the greedy algorithm share some intrinsic features, but also possess differences that make them complementary to each other. We review current bounds before we present our results.

Zhou and Spanos (2016) motivate their study of approximate submodularity within the context of sensor placement, and they consider the marginal increase in acquired information. The local submodularity index captures the difference in information yield between adding candidate sensors collectively to the established location set of sensors and adding the candidate sensors individually.

**Definition 1.** (Zhou and Spanos 2016) For a set function \( f : 2^\Omega \to \mathbb{R} \) the local submodularity index for location set \( A \subseteq \Omega \) with candidate set \( B \subseteq \Omega \) is

\[
\phi_f(A, B) = [f(A \cup B) - f(A)] - \sum_{s \in B} [f(A \cup \{s\}) - f(A)].
\]

**Definition 2.** (Zhou and Spanos 2016) For a set function \( f : 2^\Omega \to \mathbb{R} \) the submodularity index for a location set \( S \subseteq \Omega \) and cardinality parameter \( K \) is

\[
\mathcal{I}(S, K) = \max_{A \subseteq S, B \subseteq \Omega, A \cap B = \emptyset, 2 \leq |B| \leq K} \phi_f(A, B). \tag{1}
\]

If this optimization problem (1) is infeasible (i.e., no such \((A, B)\) exist), then \( \mathcal{I}(S, K) = 0 \).

Note that we have added the condition \( 2 \leq |B| \) in Definition (2) which is only implicitly included in Zhou and Spanos (2016). In addition, the following provides justification for the last line of Definition (2). The empty set is always a subset of \( S \); hence, an empty set of arguments must come from the absence of an eligible \( B \). This occurs if, for any given \( A \subseteq S \), there does not exist a set \( B \) with \( 2 \leq |B| \leq K \). Using any \( B \) with \( |B| \leq 1 \) yields \( \phi_f(A, B) = 0 \).

**Proposition 1.** (Zhou and Spanos 2016) Let \( K \) be the cardinality parameter and the number of iterations run by the greedy algorithm. Suppose \( f(\cdot) \) is a nonnegative, increasing set function, and \( \mathcal{I}(\hat{S}_K, K) \in (0, f(\hat{S}_K)) \). Then

\[
f(S_K) \geq \left(1 - \frac{1}{e} - \frac{\mathcal{I}(S_K, K)}{f(S_K)}\right) f(\hat{S}_K).
\]
Unfortunately, this bound is not well-defined in the simple case when the function evaluated at the greedy set is zero. Moreover, the authors’ original result requires the size of the greedy algorithm’s set to be equal to the cardinality parameter. Additionally, Zhou and Spanos (2016) acknowledge that computing the submodularity index exactly is hard; although they provide bounds for the submodularity index specific to their application, no general bounds on the submodularity index are given in the paper. We do note that Zhou and Spanos (2016) consider submodular function maximization \( \mathcal{I}(\mathcal{S}, L) \leq 0 \) for all \( \mathcal{S} \subseteq \Omega, L \in \{0, ..., |\Omega|\} \), but we primarily focus on non-submodular optimization in this study.

Das and Kempe (2011) define a \textit{submodularity ratio} to quantify a function’s distance from submodularity.

**Definition 3.** (Das and Kempe 2011) Let \( f(\cdot) \) be a nonnegative set function. The submodularity ratio of \( f(\cdot) \) with respect to a set \( \mathcal{S} \) and cardinality parameter \( K \geq 1 \) is

\[
\gamma(\mathcal{S}, K) = \min_{\mathcal{A} \subseteq \mathcal{S}, \mathcal{B} \subseteq \Omega, \mathcal{A} \cap \mathcal{B} = \emptyset, |\mathcal{B}| \leq K} \frac{\sum_{s \in \mathcal{B}} f(\mathcal{A} \cup \{s\}) - f(\mathcal{A})}{f(\mathcal{A} \cup \mathcal{B}) - f(\mathcal{A})}.
\]

**Proposition 2.** (Das and Kempe 2011) Let \( K \) be the cardinality parameter and \( L \) be the number of iterations run by the greedy algorithm. If \( f(\cdot) \) is a nonnegative, increasing set function, then

\[
f(\mathcal{S}_L) \geq (1 - e^{-\gamma(\mathcal{S}_L, K)})f(\hat{\mathcal{S}}_K).
\]

The submodularity ratio is not well-defined in some simple cases, which makes the bound in Proposition 2 not well-defined. A simple example is if \( f(\cdot) \) is a nonnegative, constant function. In addition, because \( |\emptyset| \leq k \), for all \( k \geq 1 \), \( \mathcal{B} = \emptyset \), one must consider \( \frac{\sum_{s \in \emptyset} f(\mathcal{A} \cup \{s\}) - f(\mathcal{A})}{f(\mathcal{A} \cup \emptyset) - f(\mathcal{A})} = 0/0 \). For another example, suppose for some \( \ell \in \{1, ..., K\} \), where \( \mathcal{S}_\ell \) is the set obtained by the greedy algorithm through \( \ell \) iterations, there exists \( \hat{\mathcal{B}} \subset \Omega \setminus \mathcal{S}_\ell \) with \( 1 \leq |\hat{\mathcal{B}}| \leq K \) and \( f(\mathcal{S}_\ell) = f(\mathcal{S}_\ell \cup \hat{\mathcal{B}}) \). Then the ratio is again undefined. The proof for the accompanying performance bound (Das and Kempe 2011, Theorem 3.2) may need to be adapted to address these undefined terms. Nevertheless, Das and Kempe (2011) do provide a multiplicative bound, which is of complementary use to additive bounds.

Horel and Singer (2016) also consider a multiplicative bound that is global, that is to say, it incorporates deviations from submodularity over the entire domain of the function. Horel and Singer (2016) state that a set function \( f(\cdot) \) is \( \epsilon \)-\textit{approximately submodular} if there exists a submodular function \( F \) such that for any \( \mathcal{S} \subseteq \Omega \), \( (1 - \epsilon)F(\mathcal{S}) \leq f(\mathcal{S}) \leq (1 + \epsilon)F(\mathcal{S}) \). Their version of approximate submodularity yields the following bound.

**Proposition 3.** (Horel and Singer 2016) Let \( K \) be the cardinality parameter and the number of iterations run by the greedy algorithm, and consider \( \epsilon \in (0, 1) \). If \( f(\cdot) \) is a nonnegative, increasing, \( \epsilon \)-approximately
submodular set function, then

$$f(S_K) \geq \frac{(1 - \epsilon)^2 f(\hat{S}_K)}{4K\epsilon + (1 - \epsilon)^2} \left( 1 - \left( \frac{(K - 1)(1 - \epsilon)^2}{K(1 + \epsilon)^2} \right)^K \right).$$

Proposition 3 is already well-defined; however, it can be localized so that less information can provide performance guarantees for the greedy algorithm. Further, the results can be extended to cases in which the cardinality parameter does not equal the number of iterations (see Section 2.3).

### 2.2 Proposed Bounds and Global-Local Trade-off

Based on what we observe from previous efforts to produce bounds for the greedy algorithm, we propose guidelines to aid the derivation of future greedy algorithm performance bounds. We propose that any such performance guarantee adheres to the following criteria: (1) it is well-defined everywhere, (2) its result holds when the number of iterations does not equal the cardinality constraint, and (3) it is amenable to considering different levels of local information, which may lead to fewer computations. There is obvious value in (1). In addition, (2) allows for flexibility in the type of approximate solution found by the greedy algorithm. If one terminates the greedy algorithm in fewer iterations than the cardinality constraint, one obtains a sparse solution. On the other hand, because the greedy algorithm is relatively inexpensive compared to an exhaustive search, one can obtain a fast, dense solution by running the algorithm for additional iterations beyond the cardinality constraint. As we show in Section 2.2, (3) enables one to seek a balance between faster bound computation and global information about the function.

In this section, we use a framework, guided by the above criteria, to propose metrics for a function’s proximity to submodularity. We also use the framework and metrics to derive new greedy algorithm bounds using global and local information. The metrics used in this section are termed the pairwise violations, and are inspired by one of the many characterizations of increasing, submodular functions.

**Proposition 4.** [Edmonds 1970] Let $f : 2^\Omega \to \mathbb{R}$. Then $f(\cdot)$ is increasing and submodular if and only if for any $A, B \subseteq \Omega, s \in \Omega \backslash A$, $f(A \cup B \cup \{s\}) - f(A \cup B) \leq f(A \cup \{s\}) - f(A)$.

We note that Edmonds (1970) provides a characterization that is equivalent to Proposition 4, though the presentation is slightly different.

To construct our bounds in accordance with the proposed criteria, we first present a metric that gives global information as to how far a function is from submodular. The pairwise violation, the first step in defining the global metric, is the worst-case violation of the condition in Proposition 4 given $A$ and $B$ with fixed cardinalities.

**Definition 4.** Let $f : 2^\Omega \to \mathbb{R}$. Consider $\ell \in \{0, \ldots, |\Omega| - 1\}, k \in \{0, \ldots, |\Omega|\}$. The $(\ell, k)$-pairwise
violation of $f(\cdot)$ is

$$d(\ell, k) = \max_{A, B \subseteq \Omega, s \in \Omega, A \cap B = \emptyset, |A| = \ell, |B| = k} f(A \cup B \cup \{s\}) - f(A \cup B)$$

$$- f(A \cup \{s\}) + f(A).$$

We return to the sensor example from Section 2.1 to provide intuition for Definition 4. The $(\ell, k)$-pairwise violation captures the case in which a single sensor added to a sparse sensor network (given by $A$) creates a smaller marginal increase in information than when the same sensor is added to a denser network ($A \cup B$).

Note that $d(\ell, k) \leq 0$ for all $\ell$ and $k$ if and only if $f(\cdot)$ is increasing and submodular. Next, we let $\delta(\ell, K) = \sum_{k=0}^{K-1} d(\ell, k)$ and define the submodularity violation as follows.

**Definition 5.** We define the $(L, K)$-submodularity violation of a function $f : 2^\Omega \to \mathbb{R}$ as

$$\Delta(L, K) = \max_{\ell \in \{0, \ldots, L\}} \delta(\ell, K),$$

for $L \in \{0, \ldots, |\Omega| - 1\}$, $K \in \{1, \ldots, |\Omega|\}$.

The submodularity violation metric can be interpreted as a sum of worst-case pairwise violations given the parameters $L$ and $K$. Note that this metric can be negative in some cases.

We derive a performance bound for the greedy algorithm at iteration $L$ given cardinality parameter $K$ as follows.

**Theorem 1.** Let $K \in \{1, \ldots, |\Omega|\}$ be the cardinality parameter and $L \in \{0, \ldots, |\Omega| - 1\}$ be the number of iterations run by the greedy algorithm. If $f(\cdot)$ is a nonnegative, increasing set function, then

$$f(S_L) \geq f(\hat{S}_K) \left[ 1 - \left(1 - \frac{1}{K}\right)^L \right],$$

for all $L \in \{0, \ldots, |\Omega| - 1\}$, $K \in \{1, \ldots, |\Omega|\}$.

Theorem [1] shows that in general, the lower bound guaranteed by the greedy algorithm has both proportional and constant components, the latter of which accounts for the submodularity violation metric (i.e., correction to the bound due to the violation of submodularity). Because $L$ corresponds to the iteration of the greedy algorithm, we need only provide a bound for $L \in \{0, \ldots, |\Omega| - 1\}$ as trivially, $f(S_{|\Omega|}) = f(\Omega)$. The min operator is only necessary for functions that are quite far from submodular. Indeed, none of the numerical examples in this paper require the min operation, and $\Delta(L, K)$ alone could equivalently replace this term in these instances. Note that if $f(\cdot)$ is nonnegative, increasing and submodular, then the submodularity violations are always nonpositive. Thus, by using the fact that $\Delta(L, K)$ is bounded above by 0 in this case, we can state the classical greedy algorithm bound from [Nemhauser et al. 1978] as a corollary of Theorem [1].

**Corollary 1.** (Nemhauser et al. 1978) Let $K \in \{1, \ldots, |\Omega|\}$ be the cardinality parameter and $L \in \{0, \ldots, |\Omega| - 1\}$ be the number of iterations run by the greedy algorithm. If $f(\cdot)$ is a nonnegative, increasing, submodular set function, then

$$f(S_L) \geq f(\hat{S}_K) \left[ 1 - (1 - 1/K)^L \right].$$
While Corollary 1 implies that the proposed bound specializes to the classical bound for a submodular function, Theorem 1 further suggests that an improvement over the classical bound is achievable for submodular functions that have negative submodularity violation (i.e., $\Delta(L, K) < 0$).

**Theorem 2.** The bound in Theorem 1 is tight.

Although it is good to know the global behavior of a function, many algorithms only call for function evaluations within a subset of the domain. Thus, we generalize the bound in Theorem 1 by considering a local bound that incorporates varying levels of localized information. For a subset $A \subseteq \Omega$, we define $\hat{d}(A, k) = \max_{B \subseteq \Omega, |B| = k, s \in \Omega \setminus A} f(A \cup B \cup \{s\}) - f(A \cup B) - f(A \cup \{s\}) + f(A)$ and refer to this quantity as the $A$-k pairwise submodularity violation. Analogously, we define $\hat{\delta}(A, K) = \sum_{k=0}^{K-1} \hat{d}(A, k)$. In addition, for a collection $\mathcal{C}$ of subsets of $\Omega$, $\hat{\Delta}(\mathcal{C}, K) = \max_{A \in \mathcal{C}} \hat{\delta}(A, K)$. Hence, $(\hat{d}(\cdot, \cdot), \hat{\delta}(\cdot, \cdot), \hat{\Delta}(\cdot, \cdot))$ are localized versions of $(d(\cdot, \cdot), \delta(\cdot, \cdot), \Delta(\cdot, \cdot))$. Note that $\hat{d}(A, k) \leq d(|A|, k)$, and $\hat{\delta}(A, K) \leq \delta(|A|, K)$, for all $A, S \subseteq \Omega, k \in \{0, ..., |\Omega| - 1\}, K \in \{1, ..., |\Omega|\}$. Moreover, if $\max |A| = L$, then $\hat{\Delta}(\mathcal{C}, K) \leq \Delta(L, K)$.

**Theorem 3.** Let $K \in \{1, ..., |\Omega|\}$ be the cardinality parameter and $L \in \{0, ..., |\Omega| - 1\}$ be the number of iterations run by the greedy algorithm. If $f(\cdot)$ is a nonnegative, increasing set function, then

$$f(S_L) \geq \left[ f(\hat{S}_K) - \min\{\hat{\Delta}(\mathcal{C}_L, K), f(\hat{S}_K)\} \right] \left[ 1 - (1 - 1/K)^L \right],$$

where $\mathcal{C}_L$ is the collection of subsets made by the greedy algorithm at each iteration.

**Theorem 4.** The bound in Theorem 3 is tight.

The proofs of Theorem 3 and Theorem 4 are similar to those of Theorem 1 and Theorem 2 respectively, and are omitted.

There are trade-offs between $\Delta(L, K)$ and $\hat{\Delta}(\mathcal{C}_L, K)$. The global metric $\Delta(L, K)$ gives insight into how close a function is to having a useful property, and it may be of use with other applications outside of greedy algorithms. However, it is more expensive to compute than $\hat{\Delta}(\mathcal{C}_L, K)$. On the other hand, the local metric $\hat{\Delta}(\mathcal{C}_L, K)$ provides the necessary information to derive a bound for the greedy algorithm, but because of its local nature, it may be less applicable in other uses. The following remark states that the local bounds can be viewed as a generalization of the global bounds. This idea can be applied to bounds in the literature.

**Remark 1.** Let $K$ be the cardinality parameter, $L$ the number of iterations run by the greedy algorithm, and $\mathcal{C} = \{S \subseteq \Omega \mid |S| \leq L\}$. Then $\hat{\Delta}(\mathcal{C}, K) = \Delta(L, K)$.  


2.3 GeneralizingExisting Bounds

As we have shown, the bounds in Theorem 1 and Theorem 3 demonstrate that one can produce nontrivial, well-defined bounds for non-submodular functions that generalize to cases in which the number of greedy iterations does not equal the cardinality parameter and capitalize on the trade-off between local and global information. In this section, we generalize bounds in the literature in the same manner, thus providing many complementary bounds to be used in varying applications to determine optimization problems in which the greedy algorithm performs well. We note that Bai and Bilmes (2018) also provide bounds for non-submodular functions that are the sum of a submodular and supermodular function. However, we do not include this work in our study as it is applicable to a more specific set of optimization problems.

First, we consider the submodularity ratio bound by Das and Kempe (2011) and provide a well-defined alternative to the original ratio.

Definition 6. Let \( f(\cdot) \) be a nonnegative set function. We define the submodularity ratio of \( f(\cdot) \) with respect to a set \( S \) and cardinality parameter \( K \geq 1 \) as

\[
\hat{\gamma}(S, K) = \max \left\{ \gamma \left| \frac{f(A \cup B) - f(A)}{f(A \cup \{s\}) - f(A)}, \forall A \subseteq S, |B| \leq k, A \cap B = \emptyset \right. \right\}.
\]

Here, when the maximum is taken over an empty set of arguments, we define its value to be \(-\infty\). It is easy to see that when \( \gamma(S, K) \) is well-defined, \( \hat{\gamma}(S, K) = \gamma(S, K) \).

Proposition 5. Let \( K \) be the cardinality parameter and \( L \) be the number of iterations run by the greedy algorithm. If \( f(\cdot) \) is a nonnegative, increasing set function, then

\[
f(S_L) \geq (1 - e^{-\hat{\gamma}(S_L, K)})f(S_K).
\]

Horel and Singer (2016) consider a multiplicative bound, but there is no stated local analog to their global bound, or to cases when \( L \neq K \). Horel and Singer (2016) state that a set function \( f(\cdot) \) is \( \epsilon \)-approximately submodular if there exists a submodular function \( F(\cdot) \) such that for any \( S \subseteq \Omega \), \( (1 - \epsilon)F(S) \leq f(S) \leq (1 + \epsilon)F(S) \). For any \( S \subseteq \Omega \), we define \( \Omega_S = \{A \subseteq \Omega | 1 \geq |A \setminus S|\} \).

Definition 7. A set function \( f : 2^\Omega \rightarrow \mathbb{R} \) is \( (S, \epsilon) \)-approximately submodular for some \( S \subseteq \Omega \) if \( f(\cdot) \) is \( \epsilon \)-approximately submodular over \( \Omega_S \).

Proposition 6. Let \( K \) be the cardinality parameter and \( L \) be the number of iterations run by the greedy algorithm. Consider \( \epsilon \in (0, 1) \). If \( f(\cdot) \) is a nonnegative, increasing, \( (S_L \cup \hat{S}_K, \epsilon) \)-approximately submodular set function, then

\[
f(S_L) \geq \frac{(1 - \epsilon)^2f(\hat{S}_K)}{4K\epsilon + (1 - \epsilon)^2} \left( 1 - \frac{(K - 1)(1 - \epsilon)^2}{K(1 + \epsilon)^2} \right)^L.
\]
We now derive a new bound that has roots in the work of Zhou and Spanos (2016) by further localizing the submodularity index. This results in a bound that requires a smaller number of function calls. The reduction in function calls is inspired by the same trade-off recognized in Section 2.2 between global and local bounds. Further, our bound is valid even when the size of the set produced by the greedy algorithm differs from the cardinality parameter. We use the bounds derived in the Section 2.2 to show our new bound is tight.

Zhou and Spanos (2016) derive a bound by considering the local submodularity index for all location sets \( S \) that are subsets of the output of the greedy algorithm. One may suspect that this approach is already localized sufficiently. However, the observations made in the previous section motivate a further reduction of the arguments considered for the local submodularity index.

**Definition 8.** For a set function \( f : 2^\Omega \to \mathbb{R} \), the (localized) submodularity indicator for a collection \( C \) of subsets of \( \Omega \) with cardinality parameter \( K \) is

\[
\hat{I}(C, K) = \max_{A \in C, B \subseteq \Omega, B \cap A = \emptyset, 2 \leq |B| \leq K} \phi_f(A, B).
\]

If this optimization problem (2) is infeasible (i.e., no such \((A, B)\) exist), then \( \hat{I}(C, K) = 0 \).

The justification for the last line of Definition 8 is similar to that of Definition 2.

**Lemma 1.** Let \( C \) be a collection of subsets of \( S \subseteq \Omega \), where \( \emptyset \in C \), and \( K \in \{0, \ldots, |\Omega| - 1\} \) is the cardinality parameter. Then \( \hat{I}(C, K) \leq I(S, K) \).

The proof of Lemma 1 is immediate from the fact that \( I(\cdot, \cdot) \) considers a superset of the \((A, B)\) pairs of \( \hat{I}(\cdot, \cdot) \).

In addition to the result of Lemma 1, the number of function calls to compute the submodularity indicator is (often strictly) less than the required function calls for the submodularity index of Zhou and Spanos (2016). In this sense, we have further localized the requirement for approximate submodularity, which helps produce a tighter bound (Theorem 5). We let \( C_L = \{\emptyset, S_1, \ldots, S_L\} \), the collection of sets chosen by the greedy algorithm through iteration \( L \).

**Theorem 5.** Let \( K \) be the cardinality parameter and \( L \) be the number of iterations run by the greedy algorithm. If \( f(\cdot) \) is a nonnegative, increasing set function, then

\[
f(S_L) \geq \min \left\{ f(S_K), \left(1 - \frac{1}{K}\right)^L \left[f(S_K) - \min\{\hat{I}(C_L, K), f(S_K)\}\right]\right\}.
\]

**Proposition 7.** Assume the conditions of Proposition 1 are satisfied, and \( f(S_K) > 0 \). Then the bound from Theorem 5 is tighter than the bound from Proposition 1.
From Proposition 7, one observes that for functions with positive submodularity indices, Theorem 5 is an improvement on Proposition 1. Also, one can use the results from Section 2.2 to prove that Theorem 5 yields a tight bound.

**Lemma 2.** Let $K$ be the cardinality parameter, and $A \subseteq \Omega$, with $|A| = \ell$. Then $\phi_f(A, B) \leq \hat{\delta}(A, K)$, for any $A \subseteq \Omega$, with $|B| = K$.

**Corollary 2.** Let $K$ be the cardinality parameter, and $A \subseteq \Omega$, with $|A| = \ell$. Then $\hat{I}(\{A\}, K) \leq \hat{\Delta}(\{A\}, K)$. Further, if $\mathcal{C} = \{A_1, \ldots, A_M\}$, then $\hat{I}(\mathcal{C}, K) \leq \hat{\Delta}(\mathcal{C}, K)$. In addition, if $L = \max_{m \in \{1, \ldots, M\}} |A_m|$, then $\hat{I}(\mathcal{C}, K) \leq \Delta(L, K)$.

**Corollary 3.** The bound in Theorem 5 is tight.

The results of Sections 2.2 and 2.3 show that the existing bounds for non-submodular functions are complementary, and that they can adhere to a common, broad, and useful set of criteria. Whether one bound is more suitable than another depends on if one wishes to establish local versus global approximate submodularity as well as a multiplicative versus additive bound.

We emphasize that all of our new bounds (Theorem 1, Theorem 3, and Theorem 5) for the greedy algorithm are for increasing, nonnegative set functions. The function’s distance to submodularity, by any of the above metrics, can be arbitrarily large or small, which can make the bounds more or less useful. This observation is in line with previous research in the area (see [Das and Kempe 2018](#) Remark 7).

### 3 Illustrative Example: Facility Location

In this section, we present a generalization of the well-known facility location problem. The objective function of this generalization is not submodular in general. We show that the pairwise violations of the problems can be bounded by exploiting the problem structure and that the objective function’s proximity to submodularity is influenced by certain problem parameters. We compute bounds from the literature and a selection of our proposed bounds.

The objective function of the uncapacitated facility location problem (UFLP) provides an example of a submodular function. An instance of UFLP is defined by $m$ facility locations ($\Omega = \{1, \ldots, m\}$), $n$ clients, demands $b \in \mathbb{R}_n^+$, fixed costs $w \in \mathbb{R}_m^+$, and facility-client revenues $v \in \mathbb{R}^{m \times n}$. We consider instances in which $v$ is nonnegative. Additionally, we assume that $w = 0$ so that the firm only assigns facilities to clients based on the variable revenue. Let $f : 2^\Omega \to \mathbb{R}$ be the objective function of the UFLP, defined by

$$f(S) = \begin{cases} 
\sum_{j=1}^n b_j \max_{i \in S} v_{ij}, & \text{if } S \neq \emptyset \\
0, & \text{if } S = \emptyset.
\end{cases}$$

---

1We refer the reader to Mirchandani and Francis (1990) for a detailed overview of facility location problems.
Here, $\mathcal{S}$ is a subset of facility locations. Under these conditions, $f(\cdot)$ is nonnegative, increasing, and submodular. We write the UFLP formulation for a set $\mathcal{A} \subseteq \Omega$ as: $\max_{\mathcal{S} \subseteq \mathcal{A}} f(\mathcal{S})$.

We consider a generalization of UFLP where the objective function is approximately submodular function in general. We introduce a nonnegative reward $u_{pq}$ associated with the simultaneous selection of two facilities $p$ and $q$, where $(p,q) \in \Omega^2$. We assume that $u_{pp} = 0$ for all $p \in \Omega$. Let $h : 2^\Omega \to \mathbb{R}$ be defined by

$$h(\mathcal{S}) = \begin{cases} \sum_{j=1}^n b_j \max_{i \in \mathcal{S}} v_{ij} + \sum_{(p,q) \in \mathcal{S}} u_{pq}, & \text{if } \mathcal{S} \neq \emptyset \\ 0 & \text{if } \mathcal{S} = \emptyset. \end{cases}$$

Then the cooperative uncapacitated facility location problem (CUFLP) over set $\mathcal{A} \subseteq \Omega$ is written as: $\max_{\mathcal{S} \subseteq \mathcal{A}} h(\mathcal{S})$.

**Remark 2.** The objective function of CUFLP is not submodular in general.

It is well-known that UFLP is NP-hard [Cornuèjols et al. (1983)]; thus, CUFLP (which includes UFLP) is NP-hard.

**Proposition 8.** Consider an instance of CUFLP. Then $d(\ell,k) \leq \left( \min\{m,\ell+k+1\} \right) \max\{u_{pq} \mid (p,q) \in \Omega^2\}$ for all $\ell \in \{0,\ldots,m-1\}$, $k \in \{0,\ldots,m\}$.

The bound on the pairwise violations in Proposition 8 can also be used to provide weaker bounds than those of Theorem 1, Theorem 3, and Theorem 5. Again, the quality of these bounds depends on the objective function’s deviation from submodularity (i.e., the cooperative bonuses) and if $\ell \ll m$ and $k \ll m$.

To demonstrate the proposed bounds, we consider a numerical example adapted from [Cornuèjols et al. (1977)] in which there are seven facilities, 12 clients, and cooperative bonuses $u_{6,7} = 25$ and $u_{pq} = 0$ otherwise. The fixed costs are set to zero, which implies that the objective function is nonnegative and increasing. We scale the cooperative bonus by $\frac{1}{t}$ where $t \in \left\{ \frac{1}{4}, \frac{1}{2}, 2 \right\}$ to generate instances with various levels of submodularity violation. We denote the resulting objective function by $h^t(\cdot)$. As $t$ increases, $\frac{1}{t} u$ decreases and $h^t(\cdot)$ approaches the submodular function $f(\cdot)$.

Figure 1 compares the optimal objective value (orange triangles), the objective value of the set chosen by the greedy algorithm (blue rectangles), and various bounds for non-submodular functions. These include our two proposed bounds, which are the global Delta Bound (Theorem 1, green crosses) and the Indicator Bound (Theorem 5, pink diamonds). The localized version of the Delta Bound is not included as it is guaranteed to lie above the global Delta Bound and below the Indicator Bound. We also include bounds from [Zhou and Spanos (2016)], [Das and Kempe (2011)], and [Horel and Singer (2016)]. The Horel and Singer (2016) bound was produced using Proposition 9.

11
(a) As \( t \) is small, the objective is far from submodular. The (global) Delta Bound does not provide a useful guarantee, but the (local) Indicator Bound does.

(b) As \( t \) increases, submodularity violation decreases and thus the adjustment in the Delta Bound also decreases.

(c) The gap between the bounds shrinks as the function approaches submodularity.

(d) When \( t = 2 \), the submodularity violation is very small, and thus the proposed bound is very close to the bound for a submodular function.

Figure 1: Greedy algorithm bounds for CFLP. The green curve is the greedy algorithm bound from Theorem 1 and the hashed region is the adjustment from the classical bound due to the submodularity violation (\( \Delta(\cdot, \cdot) \)). The pink curve is the bound from Theorem 5.
**Proposition 9.** For the above instances of cooperative facility location problems, the smallest valid $\epsilon_H$ in Proposition 6 is $u_{6,7}/f(6,7)$.

When $t$ is small, $h^t(\cdot)$ is far from submodular, in a global sense, and thus $\Delta(L, L)$ is large, which implies that the performance guarantee of the greedy algorithm may be low. As $t$ increases (e.g., from Figure 1a to Figure 1d and further towards $\infty$), $h^t(\cdot)$ approaches $f(\cdot)$ and $\Delta(L, L)$ decreases, which indicates that the greedy algorithm can perform reasonably well. The adjustments in the bounds due to non-submodularity are necessary for all $t$ considered.

The Delta Bound and the Horel and Singer (2016) bound incorporate global information, although the former is additive while the latter is multiplicative. Still, both of these bounds are more conservative than the other local behavior bounds, generally. The Delta Bound is always above the Horel and Singer (2016) bound in these examples; whether this holds true for other optimization problems remains an open question.

The Indicator Bound is often the tightest bound, or close to the best in these examples. The results reinforce Proposition 7, the Indicator Bound is always tighter than the Zhou and Spanos (2016) bound. Although there is a slight decrease in the Indicator Bound when the set size is large, this is also the case with the Horel and Singer (2016) bound. The Das and Kempe (2011) bound performs better, comparatively, on larger set sizes.

### 4 Conclusion

The greedy algorithm is a simple heuristic that is used for a wide variety of combinatorial problems. When a greedy approach is successful without known theoretical guarantees, the metrics presented in this paper can elucidate the structures behind good algorithmic performance. These generalized performance guarantees for approximately submodular set functions, along with bounds in the literature, can encourage more use of the greedy algorithm as a cheap approximation algorithm. We observe that a number of the bounds have different features—tighter bounds for small or large solutions, multiplicative or additive bounds, global or local information—that can be useful in various contexts. We also show how to approximate our new bounds using the structure of a generalized version of the facility location problem. Future directions of this study include extending the results to problems with more complex constraints and problems whose objective functions are both approximately submodular and approximately monotonic.

### Acknowledgments

The authors would like to thank their colleagues who provided valuable input, which include David Mildebrath, Seth Brown, and Silviya Valeva.
References

W. Bai and J. Bilmes. Greed is still good: Maximizing monotone Submodular+Supermodular (BP) functions. In J. Dy and A. Krause, editors, Proceedings of the 35th International Conference on Machine Learning, volume 80 of Proceedings of Machine Learning Research, pages 304–313, Stockholmsmssan, Stockholm Sweden, 10–15 Jul 2018. PMLR.

J. Christensen, J. Marks, and S. Shieber. An empirical study of algorithms for point-feature label placement. ACM Trans. Graph., 14(3):203–232, 1995.

G. Cornuéjols, M. L. Fisher, and G. L. Nemhauser. Location of bank accounts to optimize float: An analytic study of exact and approximate algorithms. Management Science, 23(8):789–810, 1977.

G. Cornuéjols, G. Nemhauser, and L. Wolsey. The uncapacitated facility location problem. Technical report, 1983. URL "https://apps.dtic.mil/dtic/tr/fulltext/u2/a140000.pdf".

A. Das and D. Kempe. Submodular meets spectral: Greedy algorithms for subset selection, sparse approximation and dictionary selection. In Proceedings of the 28th International Conference on International Conference on Machine Learning, ICML’11, pages 1057–1064, USA, 2011. Omnipress. ISBN 978-1-4503-0619-5.

A. Das and D. Kempe. Approximate submodularity and its applications: Subset selection, sparse approximation and dictionary selection. Journal of Machine Learning Research, 19(3):1–34, 2018.

J. Edmonds. Submodular functions, matroids, and certain polyhedra. In Combinatorial Structures and Their Applications (Proceedings of Calgary International Conference on Combinatorial Structures and Their Applications), pages 69–87, New York, NY, USA, 1970. Gordon and Breach.

M. Fisher, G. Nemhauser, and L. Wolsey. An analysis of approximations for maximizing submodular set functions - II. Mathematical Programming Study, 8:73–87, 1978.

J. Gottlieb, M. Puchta, and C. Solnon. A study of greedy, local search, and ant colony optimization approaches for car sequencing problems. In S. Cagnoni, C. G. Johnson, J. J. R. Cardalda, E. Marchiori, D. W. Corne, J.-A. Meyer, J. Gottlieb, M. Middendorf, A. Guiller, G. R. Raidl, and E. Hart, editors, Applications of Evolutionary Computing, pages 246–257, Berlin, Heidelberg, 2003. Springer Berlin Heidelberg.

T. Horel and Y. Singer. Maximization of approximately submodular functions. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, Advances in Neural Information Processing Systems 29, pages 3045–3053. Curran Associates, Inc., 2016.

A. Krause, A. Singh, and C. Guestrin. Near-optimal sensor placement in Gaussian processes: Theory, efficient algorithms and empirical studies. Journal of Machine Learning, 9:235–284, 2008.

P. B. Mirchandani and R. L. Francis, editors. Discrete Location Theory. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley, 1 edition, 1990.

G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher. An analysis of approximations for maximizing submodular set functions—I. Mathematical Programming, 14(1):265–294, 1978.

P. Tuffery, F. Guyon, and P. Derreumaux. Improved greedy algorithm for protein structure reconstruction. Journal of Computational Chemistry, 26(5):506–513, 2005.
Y. Zhou and C. J. Spanos. Causal meets submodular: Subset selection with directed information. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems 29*, pages 2649–2657. Curran Associates, Inc., 2016. URL http://papers.nips.cc/paper/6384-causal-meets-submodular-subset-selection-with-directed-information.pdf
Appendix

Lemma 3. Let \( \{a_i\} \) be a sequence in \( \mathbb{R} \) such that \( a_{i+1} \leq \alpha a_i + \beta \), where \( \alpha \in \mathbb{R}_+, \beta \in \mathbb{R} \). Let \( b_0 = a_0, b_{i+1} = \alpha b_i + \beta \). Then \( b_i \geq a_i \) for all \( i \in \mathbb{N} \).

Proof: We prove by induction. Note that \( a_0 = b_0 \), and the base case of \( n = 1 \) is trivial. Assume for all \( n \leq N - 1 \), for some \( N \in \mathbb{N} \), \( a_n \leq b_n \).

\[
a_N \leq \alpha a_{N-1} + \beta \\
\leq \alpha b_{N-1} + \beta \\
= b_N.
\]

By induction, \( a_n \leq b_n \) for all \( n \in \mathbb{N} \). \( \square \)

Lemma 4. Let \( \alpha, \beta \), and \( b_0 \in \mathbb{R} \), where \( \alpha \neq 1 \). Define the sequence \( \{b_i\} \) by \( b_{i+1} = \alpha b_i + \beta \). Then \( b_i = \alpha^i (b_0 - \frac{\beta}{1-\alpha}) + \frac{\beta}{1-\alpha} \).

Proof: Let \( \tilde{b}_n = \alpha^n (b_0 - \frac{\beta}{1-\alpha}) + \frac{\beta}{1-\alpha} \). We show by induction that \( b_n = \tilde{b}_n \) for all \( n \). The base case is \( n = 0 \):

\[
\tilde{b}_0 = \alpha^0 (b_0 - \frac{\beta}{1-\alpha}) + \frac{\beta}{1-\alpha} \\
= b_0.
\]

Assume for all \( n \leq N - 1 \), \( \tilde{b}_n = b_n \). We show \( \tilde{b}_N = b_N \).

\[
\begin{align*}
b_{N-1} &= \tilde{b}_{N-1} \text{ (by the inductive hypothesis)} \\
&= \alpha^{N-1} (b_0 - \frac{\beta}{1-\alpha}) + \frac{\beta}{1-\alpha}.
\end{align*}
\]

\[
\begin{align*}
b_N &= \alpha b_{N-1} + \beta \\
&= \alpha \tilde{b}_{N-1} + \beta \\
&= \alpha \left[ \alpha^{N-1} (b_0 - \frac{\beta}{1-\alpha}) + \frac{\beta}{1-\alpha} \right] + \beta \\
&= \alpha^N (b_0 - \frac{\beta}{1-\alpha}) + \frac{\beta}{1-\alpha}.
\end{align*}
\]

This concludes the proof. \( \square \)

Lemma 5. Let \( \{a_i\} \) be a sequence in \( \mathbb{R} \) such that \( a_{i+1} \leq \alpha a_i + \beta \), for some \( \alpha \in \mathbb{R}_+, \beta \in \mathbb{R} \), where \( \alpha \neq 1 \). Let \( b_0 = a_0 \) and \( b_{i+1} = \alpha b_i + \beta \). Then \( a_i \leq \alpha^i (a_0 - \frac{\beta}{1-\alpha}) + \frac{\beta}{1-\alpha} \).
The proof of Lemma 5 follows from Lemma 3 and Lemma 4.

Define an order on the elements of $\Omega$. For $k \in \{0, \ldots, K\}$, denote the elements of $\hat{S}_K$ by $\hat{S}_K(k) = \{\hat{s}_1^K, \ldots, \hat{s}_k^K\}$ (so that $\hat{S}_K(K) = \hat{S}_K$). Denote the $\ell^{th}$ element selected by the greedy algorithm by $s_\ell$. Note that we interpret $\{\hat{s}_1^K, \ldots, \hat{s}_k^K\} = \emptyset$ when $k = 0$ and $S_0 = \emptyset$. For any $K, L \in \mathbb{Z}$ and $K \neq 0$, we let $\nu_{KL} = \left(\frac{K-1}{K}\right)^L$.

**Proof of Theorem 1**

Fix $L$ and $K$. A telescoping sum argument shows that for any $\ell \in \{0, \ldots, L\}$,

$$f(\hat{S}_K \cup S_\ell) = f(S_\ell) + \sum_{k=0}^{K-1} f(\hat{S}_K(k+1) \cup S_\ell) - \sum_{k=0}^{K-1} f(\hat{S}_K(k) \cup S_\ell).$$  \hfill (3)

Fix $k \in \{0, \ldots, K-1\}$. By the definition of $d(\ell, k)$,

$$f(\hat{S}_K(k+1) \cup S_\ell) - f(\hat{S}_K(k) \cup S_\ell) \leq f(\{\hat{s}_k^{k+1}\} \cup S_\ell) - f(S_\ell) + d(\ell, k).$$  \hfill (4)

We plug the bound obtained in (4) into (3), and use the fact that $f(S_{\ell+1}) \geq (\hat{s}_k^{k+1} \cup S_\ell)$ to obtain

$$f(\hat{S}_K \cup S_\ell) \leq f(S_\ell) + \sum_{k=0}^{K-1} [f(S_{\ell+1}) - f(S_\ell) + d(\ell, k)] \rightarrow f(\hat{S}_K \cup S_\ell) - f(S_\ell) \leq K(f(S_{\ell+1}) - f(S_\ell)) + \delta(\ell, K).$$  \hfill (5)

Because $f$ is increasing,

$$f(\hat{S}_K) - f(S_\ell) \leq K(f(S_{\ell+1}) - f(S_\ell)) + \delta(\ell, K).$$  \hfill (6)

Some additional arithmetic yields

$$f(\hat{S}_K) - f(S_{\ell+1}) \leq \nu_{K1}(f(\hat{S}_K) - f(S_\ell)) + \frac{\delta(\ell, K)}{K}.$$

By Lemma 5 and the nonnegativity of $f$, we have, for all $\ell \in \{0, \ldots, L\}$,

$$f(\hat{S}_K) - f(S_\ell) \leq \nu_{K\ell}(f(\hat{S}_K) - \Delta(L, K)) + \Delta(L, K).$$
Let \( \ell = L \). Then simple rearrangement of terms yields
\[
[f(\hat{S}_K) - \Delta(L, K)] [1 - v_{KL}] \leq f(S_L), \tag{7}
\]
which completes the proof.

\[\square\]

**Proof of Theorem 2**: We show that there exist tight examples in which \( K = L \) and \( L \in \{1, \ldots, |\Omega| - 1\} \). Fisher et al. (1978) proved that, for each such \( L \), there exists a nonnegative, increasing, submodular function \( f \) such that
\[
f(\hat{S}_L) [1 - v_{LL}] = f(S_L). \tag{8}
\]
By Theorem 1, \[
[f(\hat{S}_L) - \Delta(L, L)] [1 - v_{LL}] \leq f(S_L) = f(\hat{S}_L) [1 - v_{LL}],
\]
which implies that \( \Delta(L, L) \geq 0 \).

By construction, \( f \) is submodular; hence, for all \( s \in \Omega, A, B \subset \Omega, 0 \geq f(A \cup B \cup \{s\}) - f(A \cup B) - f(A \cup \{s\}) + f(A) \). This implies \( d(\ell, k) \leq 0 \), for all \( \ell \in \{1, \ldots, |\Omega|\}, k \in \{0, \ldots, |\Omega| - 1\} \). By the definition of \( \delta(\ell, L) \), we have \( \delta(\ell, L) = \sum_{k=0}^{L-1} d(\ell, k) \leq 0 \). This implies
\[
\Delta(L, L) = \max_{\ell \in \{1, \ldots, L\}} \delta(\ell, L) \leq 0.
\]
Thus, \( \Delta(L, L) = 0 \), and the bound from Theorem 1 is
\[
f(\hat{S}_L) [1 - v_{LL}] \leq f(S_L),
\]
which we already stated is an equality in (8).

\[\square\]

**Proof of Proposition 6**: Fix \( L \) and \( K \). Consider \( \ell \in \{1, \ldots, L - 1\} \), and let \( F : 2^{\Omega} \to \mathbb{R} \) be a function that is submodular over \( \Omega_{S_L \cup \hat{S}_K} \) and \((1 - \epsilon)F(S) \leq f(S) \leq (1 + \epsilon)F(S)\), for all \( S \in \Omega_{S_L \cup \hat{S}_K} \). By the local submodularity of \( F \), the greedy algorithm, and the approximate local submodularity of \( f \), we have
\[
F(\hat{S}_K) \leq F(S_\ell) + \sum_{s \in \hat{S}_K} [F(S_\ell \cup \{s\}) - F(S_\ell)]
\leq F(S_\ell) + \sum_{s \in \hat{S}_K} \left[ \frac{1}{1 - \epsilon} f(S_\ell \cup \{s\}) - F(S_\ell) \right]
\leq F(S_\ell) + \sum_{s \in \hat{S}_K} \left[ \frac{1}{1 - \epsilon} f(S_{\ell+1}) - F(S_\ell) \right]
\]
\[ F(S_\ell) + \sum_{s \in \hat{S}_K} \left[ \frac{1 + \epsilon}{1 - \epsilon} F(S_{\ell+1}) - F(S_\ell) \right] \leq F(S_\ell) + K \left[ \frac{1 + \epsilon}{1 - \epsilon} F(S_{\ell+1}) - F(S_\ell) \right]. \]

Rearranging the above inequality yields

\[ K \frac{1 + \epsilon}{1 - \epsilon} F(S_{\ell+1}) \geq (K - 1) F(S_\ell) + F(\hat{S}_K), \]

which is equivalent to

\[ f(S_{\ell+1}) \geq \frac{vK_1(1 - \epsilon)^2 F(S_\ell)}{(1 + \epsilon)} + \frac{(1 - \epsilon)^2 F(\hat{S}_K)}{K(1 + \epsilon)} \geq vK_1(1 - \epsilon)^2 f(S_\ell) + \frac{1}{K} \frac{(1 - \epsilon)^2}{(1 + \epsilon)^2} f(\hat{S}_K). \]

The last inequality comes from local approximate submodularity. As stated in Horel and Singer (2016), this is an inductive inequality \( a_{\ell+1} \geq \alpha a_\ell + \beta, a_0 = 0 \), from which it follows that \( a_\ell \geq \frac{\beta}{1 - \alpha} (1 - \alpha^\ell) \). Hence, we have

\[ f(S_\ell) \geq \frac{(1 - \epsilon)^2 f(\hat{S}_K)}{4K\epsilon + (1 - \epsilon)^2} \left( 1 - \left( \frac{vK_1(1 - \epsilon)^2}{(1 + \epsilon)^2} \right)^\ell \right) \]

and this implies

\[ f(S_L) \geq \frac{(1 - \epsilon)^2 f(\hat{S}_K)}{4K\epsilon + (1 - \epsilon)^2} \left( 1 - \left( \frac{vK_1(1 - \epsilon)^2}{(1 + \epsilon)^2} \right)^L \right) \]

which completes the proof. \( \square \)

**Proof of Theorem 5** Suppose \( |\hat{S}_K \setminus S_{L-1}| \geq 2 \). Fix \( \ell \in \{0, ..., L - 1\} \). Observe that

\[ K(f(S_{\ell+1}) - f(S_\ell)) = K \sum_{s \in S_{\ell+1} \setminus S_\ell} (f(S_\ell \cup \{s\}) - f(S_\ell)) \geq \sum_{s \in \hat{S}_K \setminus S_\ell} (f(S_\ell \cup \{s\}) - f(S_\ell)). \]  

By the definition of the local submodularity index (Definition 1), the right-hand side of (9) can be rewritten
as
\[-\phi_f(S_{\ell}, \hat{S}_K \setminus S_{\ell}) + f(\hat{S}_K \cup S_{\ell}) - f(S_{\ell}).\]  

(10)

Hence,
\[f(\hat{S}_K \cup S_{\ell}) - f(S_{\ell}) \leq K(f(S_{\ell+1}) - f(S_{\ell})) + \hat{I}(S_{\ell}, |\hat{S}_K \setminus S_{\ell}|).\]

The function \(f(\cdot)\) is increasing; therefore,
\[f(\hat{S}_K) - f(S_{\ell}) \leq K(f(S_{\ell+1}) - f(S_{\ell})) + \hat{I}(S_{\ell}, |\hat{S}_K \setminus S_{\ell}|),\]

which is equivalent to
\[f(\hat{S}_K) - f(S_{\ell+1}) \leq v_K(f(\hat{S}_K) - f(S_{\ell})) + \frac{\hat{I}(S_{\ell}, |\hat{S}_K \setminus S_{\ell}|)}{K}.\]

From this, it follows that
\[f(\hat{S}_K) - f(S_{\ell+1}) \leq v_K(f(\hat{S}_K) - f(S_{\ell})) + \frac{\hat{I}(S_{\ell}, K)}{K}.\]

The last inequality uses the fact that \(\hat{I}(S_{\ell}, J) \leq \hat{I}(S_{\ell}, K)\) for all \(J \leq K\) with \(|\hat{S}_K \setminus S_J| \geq 2\). By Lemma 5 and the nonnegativity of \(f\),
\[f(\hat{S}_K) - f(S_{\ell}) \leq v_K(f(\hat{S}_K) - \hat{I}(S_{\ell}, K)) + \hat{I}(S_{\ell}, K).\]

When \(\ell = L\),
\[f(\hat{S}) - f(S_L) \leq v_K\left[f(\hat{S}) - \hat{I}(S_L)\right] + \hat{I}(S_L, K)\]
\[\iff f(S_L) \geq (1 - v_K)\left[f(\hat{S}) - \hat{I}(S_L)\right]\]
\[\Rightarrow f(S_L) \geq (1 - v_K)\left[f(\hat{S}) - \min\{\hat{I}(S_L), f(\hat{S}_K)\}\right].\]

(11)

(12)

(13)

Now, suppose that \(|\hat{S}_K \setminus S_{L-1}| \leq 1\). If \(\hat{S}_K \subseteq S_{L-1}\), then by the monotonicity of \(f(\cdot)\), \(f(S_L) \geq f(\hat{S}_K)\). Otherwise, let \(\{s\} = \hat{S}_K \setminus S_L\) and \(\{t\} = S_L \setminus S_{L-1}\). By the greedy algorithm and monotonicity, \(f(S_L) = \)
\[ f(S_{L-1} \cup \{t\}) \geq f(S_{L-1} \cup \{s\}) \geq f(\hat{S}_K). \] Therefore,
\[
f(S_L) \geq \min \left\{ f(\hat{S}_K), (1 - \nu_{KL}) \left[ f(\hat{S}_K) - \lambda \right] \right\},
\]
where \( \lambda = \min\{\bar{I}(C, K), f(\hat{S}_K)\} \), and the theorem is proved. \( \square \)

Proof of Proposition 7: Observe:
\[
(1 - \nu_{KK}) \left[ f(\hat{S}_K) - \min\{\bar{I}(S_K, K), f(\hat{S}_K)\} \right]
\geq (1 - \frac{1}{e}) \left[ f(\hat{S}_K) - \min\{\bar{I}(S_K, K), f(\hat{S}_K)\} \right]
\geq \left( 1 - \frac{1}{e} - \frac{\min\{\bar{I}(S_K, K), f(\hat{S}_K)\}}{f(S_K)} \right) f(\hat{S}_K)
\geq \left( 1 - \frac{1}{e} - \frac{\bar{I}(S_K, K)}{f(\hat{S}_K)} \right) f(\hat{S}_K).
\]

Proof of Lemma 2: Let \( B(k) = \{b_1, ..., b_k\} \), for each \( k \in \{0, ..., K\} \). Observe:
\[
\phi_f(A, B) = f(A \cup B) - f(A)
- \sum_{b \in B} [f(A \cup \{b\}) - f(A)]
= \sum_{k=0}^{K-1} [f(A \cup B(k+1)) - f(A \cup B(k))]
- \sum_{k=0}^{K-1} [f(A \cup \{b_{k+1}\}) + f(A)]
\leq \sum_{k=0}^{K-1} \hat{d}(A, k)
= \delta(A, K).
\]

Proof of Proposition 8: Because \( u \) is nonnegative, \( f(S) \leq h(S) \) for all \( S \subseteq \Omega \). Further, \( f(S) \geq h(S) - \left( \min\{m, \ell + k + 1\} \right) \max\{u_{pq} \mid (p, q) \in \Omega^2\} \). Let \( A, B \subseteq \Omega, s \in \Omega \setminus A \), where \( |A| = \ell \in \{0, ..., m-1\} \) and
$|B| = k \in \{0, \ldots, m\}$.

\[
\begin{align*}
&h(A \cup B \cup \{s\}) - h(A \cup B) - h(A \cup \{s\}) + h(A) \\
&\leq h(A \cup B \cup \{s\}) - f(A \cup B) - f(A \cup \{s\}) + h(A) \\
&\leq f(A \cup B \cup \{s\}) - f(A \cup B) - f(A \cup \{s\}) + f(A) \\
&\quad + \left(\frac{\min\{m, \ell + k + 1\}}{2}\right) \max\{u_{pq} \mid (p, q) \in \Omega^2\} \\
&\leq \left(\frac{\min\{m, \ell + k + 1\}}{2}\right) \max\{u_{pq} \mid (p, q) \in \Omega^2\}.
\end{align*}
\]

It follows immediately that

\[
d(\ell, k) \leq \left(\frac{\min\{m, \ell + k + 1\}}{2}\right) \max\{u_{pq} \mid (p, q) \in \Omega^2\}.
\]

\[\square\]

**Proof of Proposition 9**: For $\epsilon$ to be valid for Proposition 6, $(1 - \epsilon)f(S) \leq h(S) \leq (1 + \epsilon)f(S)$, for all $S \subseteq \Omega$. Because $u_{pq} \geq 0$, for all $(p, q) \in \Omega^2$, $f(S) \leq h(S)$, for all $S \subseteq \Omega$. Because $\epsilon_H > 0$ and $f(\cdot)$ is nonnegative, it satisfies

\[
h(S) \geq f(S) - \epsilon_H f(S), \forall S \subseteq \Omega.
\]

Note that by the values of $u$, $\min\{h(S) \mid h(S) > f(S), S \subseteq \Omega\} = h(\{6, 7\})$, and $\max\{h(S) - f(S) \mid S \subseteq \Omega\} = u_{6,7}$. Hence, if $\{6, 7\} \not\subseteq S$, then $h(S) = f(S) \leq (1 + \epsilon_H)f(S)$. If $\{6, 7\} \subseteq S$, then

\[
\begin{align*}
h(S) &= f(S) + u_{6,7} \\
&\geq f(S) + \frac{u_{6,7}}{f(\{6, 7\})}f(S) \\
&= f(S) + \epsilon_H f(S).
\end{align*}
\]

Further, $h(\{6, 7\}) = f(\{6, 7\}) + \epsilon_H f(\{6, 7\})$; hence, $\epsilon_H$ cannot be decreased. \[\square\]