Approximate Submodularity and Its Implications in Discrete Optimization

Temitayo Ajayi *[1], Taewoo Lee †[2] and Andrew J. Schaefer ‡[1]

[1]Rice University, Department of Computational and Applied Mathematics

[2]Department of Industrial Engineering, University of Houston

Abstract. Submodularity, a discrete analog of convexity, is a key property in discrete optimization that features in the construction of valid inequalities and analysis of the greedy algorithm. In this paper, we broaden the approximate submodularity literature, which so far has largely focused on variants of greedy algorithms and iterative approaches. We define metrics that quantify approximate submodularity and use these metrics to derive properties about approximate submodularity preservation and extensions of set functions. We show that previous analyses of mixed-integer sets, such as the submodular knapsack polytope, can be extended to the approximate submodularity setting. In addition, we demonstrate that greedy algorithm bounds based on our notions of approximate submodularity are competitive with those in the literature, which we illustrate using a generalization of the uncapacitated facility location problem.

Keywords— Approximate submodularity, valid inequalities, greedy algorithm, facility location

1 Introduction

Exploiting structural properties in discrete optimization can lead to successful algorithms and heuristics. A classical property that is frequently used in discrete optimization is submodularity. Let Ω be a finite set

*ta21@rice.edu

†tlee20@central.uh.edu

‡andrew.schaefer@rice.edu

of elements, and let 2^Ω denote the power set of Ω. A set function $f : 2^\Omega \to \mathbb{R}$ is *submodular* if for any $\mathcal{A} \subseteq \mathcal{B} \subseteq \Omega$ and $s \in \Omega \backslash \mathcal{B}$, $f(\mathcal{B} \cup \{s\}) - f(\mathcal{B}) \leq f(\mathcal{A} \cup \{s\}) - f(\mathcal{A})$. For some problems, submodularity provides guarantees for solution approaches, such as the greedy algorithm. In recent years, researchers have expanded algorithm analysis to *approximately submodular* functions. However, much of the initial focus on approximate submodularity has remained within performance guarantees for algorithms. In this paper, we propose approximate submodularity metrics to study multiple implications of approximate submodularity, including the derivation of valid inequalities, properties of extensions on the unit hypercube, as well as greedy algorithm performance bounds. Our work applies to any nonnegative and monotonic set function, and our analyses often follow arguments similar to those of analogous results in the submodular context.

1.1 Extensions of Set Functions

Continuous relaxations of problems are often used as direct approximation techniques (e.g., solving the linear programming relaxation of a mixed-integer program) because they are easier to solve; frequently, these relaxations have a polynomial-time algorithm. In discrete optimization, extensions of set functions can transform discrete optimization problems into continuous optimization problems, which may have good algorithms or approximation schemes. Formally, an *extension* of a set function $f : 2^\Omega \to \mathbb{R}$ is a function $F : D \to \mathbb{R}$ such that $D \supset \mathbb{B}^{|\Omega|}$ and $F(x(\mathcal{S})) = f(\mathcal{S})$, for all $\mathcal{S} \subseteq \Omega$. Here, $x(\mathcal{S})$ denotes the characteristic vector of the set $\mathcal{S}$. We focus on extensions defined on the unit hypercube $[0, 1]^{|\Omega|}$. Notably, the *Lovász extension* (Lovász 1983) for set functions is convex if and only if the set function is submodular. Moreover, when the set function is submodular, the Lovász extension is equal to the convex closure. The convex closure is difficult to compute in general; in contrast, computing the Lovász extension is comparatively simple, which makes submodularity a valuable property when considering solution methods that use the convex closure. Another useful extension of set functions is the multilinear extension. In particular, when the set function is submodular, its multilinear extension is up-concave (Definition 14), which has prompted researchers to use it in solution approaches for maximization problems.

The Lovász and multilinear extensions are not, in general, convex and up-concave, respectively, when the underlying set function is not submodular, but there is no analogous study of properties of these extensions when the set function is approximately submodular. Researchers have considered relaxed properties of extensions in different areas of discrete convexity. For instance, after early studies of alternative definitions of discrete convexity, such as M- and L- convexity (e.g., Murota (1998), Fujishige and Murota (2000)), new research emerged on relaxed M- and L- convex analysis in the form of quasi- M- and L-convex functions (Murota and Shioura 2003).

1.2 Valid Inequalities

Valid inequalities are crucial for solving mixed-integer programs. Valid inequalities can cut off solutions to relaxations so that the new problem's feasible region more closely approximates the convex hull. Overviews of well-known techniques for deriving valid inequalties in (mixed-)integer programming include Van Roy and Wolsey (1986) and Cornuéjols (2008).

A foundational problem in discrete optimization is the knapsack problem. Early studies on valid inequalities and the facial structure of the linear knapsack polytope include Balas (1975). Nonlinear knapsack problems have been studied as well (e.g., Hochbaum (1995)). Song et al. (2014) consider a chance-constrained knapsack problem and propose probabilistic covers to generate valid inequalities. Atamtürk and Narayanan (2009) study the submodular knapsack polytope, in which the constraint function is submodular. Submodular functions feature in the constraints of other optimization problems. Atamtürk and Narayanan (2008), Atamtürk and Gómez (2020), and Gómez (2020) study conic-quadratic constraints and objective functions of mixed-integer programs, which have applications in value-at-risk minimization. The authors use the fact that the function defining the constraint or objective is submodular with respect to the binary variables when viewed as a set function. Atamtürk and Narayanan (2020) study valid inequalities and outer approximations of the epigraphs of submodular and general set functions. Our study is the first to use approximate submodularity to derive valid inequalities for mixed-integer sets defined by approximately submodular functions.

1.3 Greedy Algorithm

Some of the earliest work in discrete optimization and the greedy algorithm involved optimization over matroids, an abstraction of linear independence in vector spaces introduced by Whitney (1935). Given a matroid M defined by its finite set of elements Ω and its independent sets $\mathcal{F}$, consider the problem $\max_{\mathcal{S}\subseteq\Omega}\left\{\sum_{j\in\mathcal{S}} c_j \mid \mathcal{S}\in\mathcal{F}\right\}$, where c_j is the weight assigned to element $j\in\Omega$. This is known as the maximum-weight independent set problem over a matroid and the greedy algorithm is guaranteed to find an optimal solution (Edmonds 1971, Nemhauser and Wolsey 1988). The problem can be generalized by replacing the linear weight function with a submodular function. Maximizing a submodular function over a uniform matroid is a prominent optimization problem, which is equivalent to maximizing a submodular function subject to a cardinality constraint: $\max_{\mathcal{S}\subseteq\Omega}\{f(\mathcal{S}) : |\mathcal{S}|\leq K\}$, where $K\in\{1,...,|\Omega|\}$ is the *maximum cardinality parameter*.

In a seminal paper, Cornuéjols et al. (1977) analyze the greedy algorithm for the uncapacitated facility location problem with nonnegative profits and no fixed costs, and they prove it has a multiplicative performance bound. More generally, Nemhauser et al. (1978) and Fisher et al. (1978) show that when maximizing a nonnegative, increasing, submodular function, the greedy algorithm is within $1-(1-\frac{1}{K})^K$ of optimality, where K is both the number of iterations run by the algorithm and the maximum cardinality parameter. Submodular functions have since been of interest to the operations research community (e.g., Nemhauser and Wolsey (1978), Sviridenko (2004)).

Others have studied performance bounds for the greedy algorithm without both monotonicy and submodularity. Krause et al. (2008) show that if a function is submodular and *approximately monotonic*, then the greedy algorithm has a performance bound, and researchers have studied the greedy algorithm with approximate submodularity (Das and Kempe 2011, Horel and Singer 2016, Zhou and Spanos 2016), see Section 5.1 for details. The authors' bounds are produced using different notions of approximate submodularity; new metrics with different properties can produce different bounds. Some bounds only analyze the greedy algorithm when the number of greedy algorithm iterations equals the maximum cardinality parameter, which may be too restrictive. Thus, there is a need to derive generalized and well-defined bounds for the greedy algorithm's performance when optimizing increasing, nonnegative, approximately submodular functions.

1.4 Summary of Contributions

Previous studies on approximate submodularity have focused primarily on performance bounds for greedy algorithms and other iterative selection approaches (see Section 1.3). We provide other applications of approximate submodularity that yield insights in new areas. New methodological applications are still emerging, even outside of the greedy algorithm, in which approximate submodularity can be used to extend results that depend on submodularity.

Hence, in this paper, we leverage approximate submodularity to generalize results in multiple discrete optimization areas. Our contributions are as follows:

1. In Section 2, we introduce novel approximate submodularity metrics. Our study reveals fundamental properties about our metrics, which we use to show which operations preserve approximate submodularity, with respect to the metrics. Using our metrics, in Section 3, we derive results for extensions of approximately submodular functions. We show that the proposed metrics can be used to (1) prove the Lovász extension is approximately convex, and (2) guarantee that the multilinear extension is approximately up-concave.

2. We study mixed-integer sets defined by approximately submodular functions in Section 4. We use the proposed metrics to adapt analogous analyses for the submodular setting, thus deriving new valid inequalities for cases when the set function is approximately submodular. We demonstrate that by using these valid inequalities for an approximately submodular knapsack polytope, fractional solutions can be cut off and an integral optimal solution can be obtained.

3. By utilizing the proposed metrics, we construct multiple greedy algorithm bounds for approximately

submodular functions (Section 5). We prove that these novel bounds are tight, and we compare these bounds computationally to those in the literature using a modified uncapacitated facility location problem. Our bounds are on par with, or improve upon, those in the literature.

We note that there are several cases in which our proofs are similar to those of analogous results in the submodular setting. Our primary message is that in a broad set of areas in discrete optimization, one can use approximate submodularity to generalize both classical and more recent results.

2 Approximate Submodularity Metrics

In this section, we discuss various approximate submodularity metrics. We use the term "metrics" loosely, as some of the metrics are not subadditive and none are positive definite, both of which are part of the formal definition of a metric. However, the approximate submodularity metrics we discuss all indicate a notion of distance to submodularity. Throughout this work, if ζ is an approximate submodularity metric, $\zeta[f]$ is the metric value for f. We review some existing approximate submodularity metrics in Section 2.1. In Section 2.2, we propose our metrics used throughout this paper. In Section 2.3, we prove some properties of our metrics, including that they are sublinear, and we show how they relate to the preservation of approximate submodularity.

2.1 Existing Notions of Approximate Submodularity

The approximate submodularity literature has grown quickly over the last few years (e.g, Borodin et al. (2014), Bian et al. (2017), and Bai and Bilmes (2018) and the references therein). We review a selection of studies that address approximate submodularity in different ways. Zhou and Spanos (2016) motivate their study of approximate submodularity within the context of sensor placement and consider the marginal increase in acquired information. They define the local submodularity index to capture the difference in information yield between adding candidate sensors collectively to the established set of sensors and adding the candidate sensors individually. The authors use the local submodularity index to define their submodularity index, which they use to analyze the greedy algorithm and a randomized variant.

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

$$\phi^{\mathcal{A},\mathcal{B}}[f] \coloneqq [f(\mathcal{A} \cup \mathcal{B}) - f(\mathcal{A})] - \sum_{s \in \mathcal{B}} [f(\mathcal{A} \cup \{s\}) - f(\mathcal{A})] \,.$$

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

$$\mathcal{I}^{\mathcal{S},K}[f] := \max_{\substack{\mathcal{A}\subseteq\mathcal{S},\mathcal{B}\subseteq\Omega,\\ \mathcal{A}\cap\mathcal{B}=\emptyset,2\leq|\mathcal{B}|\leq K}} \phi^{\mathcal{A},\mathcal{B}}[f]. \tag{1}$$

If no such $(\mathcal{A},\mathcal{B})$ *exist in* (1)*, then* $\mathcal{I}^{\mathcal{S},K}[f] = 0$.

The following provides justification for the last line of Definition 2. The empty set is always a subset of $\mathcal{S}$; hence, an empty set of arguments must come from the absence of an eligible $\mathcal{B}$. This occurs if, for any given $\mathcal{A} \subseteq \mathcal{S}$, there does not exist a set $\mathcal{B}$ with $2 \leq |\mathcal{B}| \leq K$. Using any $\mathcal{B}$ with $|\mathcal{B}| \leq 1$ yields $\phi^{\mathcal{A},\mathcal{B}}[f] = 0$.

Das and Kempe (2011) define a submodularity ratio to quantify the degree to which a function violates submodularity, and we present a modified version of their metric.

Definition 3. *Let* f *be a nonnegative set function. Define the* submodularity ratio *of* f *with respect to a set* $\mathcal{S}$ *and maximum cardinality* $K \geq 1$ *as*

$$\widehat{\gamma}^{\mathcal{S},K}[f] := \max\{\gamma \mid [f(\mathcal{A}\cup\mathcal{B}) - f(\mathcal{A})]\gamma \leq \sum_{s\in\mathcal{B}} f(\mathcal{A}\cup\{s\}) - f(\mathcal{A}), \forall\ \mathcal{A}\subseteq\mathcal{S}, |\mathcal{B}|\leq K, \mathcal{A}\cap\mathcal{B}=\emptyset\}.$$

Here, when the maximum is taken over the empty set, we define its value to be $-\infty$. It is easy to see that wherever $\gamma^{\mathcal{S},K}[f]$, from Das and Kempe (2011), is well-defined, $\widehat{\gamma}^{\mathcal{S},K}[f] = \gamma^{\mathcal{S},K}[f]$.

Horel and Singer (2016) use a notion of approximate submodularity that is directly tied to the existence of a nearby submodular function. Specifically, Horel and Singer (2016) define a set function f to be ϵ-*approximately submodular* if there exists a submodular function F such that for any $\mathcal{S} \subseteq \Omega$, $(1-\epsilon)F(\mathcal{S}) \leq f(\mathcal{S}) \leq (1+\epsilon)F(\mathcal{S})$.

2.2 Proposed Notions of Approximate Submodularity

The metrics reviewed in Section 2.1 are all motivated by analyzing the greedy algorithm and its variants. However, notions of approximate submodularity can have utility outside of this specific use case. We propose new metrics, some of which are designed for the greedy algorithm like the existing ones in Section 2.1, whereas others are for general approximately submodular analysis.

We begin with the most general (global) metric. It is inspired directly from an equivalent definition of submodularity: $f(\mathcal{A}\cup\mathcal{B}) + f(\mathcal{A}\cap\mathcal{B}) \leq f(\mathcal{A}) + f(\mathcal{B})$, for all $\mathcal{A},\mathcal{B} \subseteq \Omega$.

Definition 4. *Let* $f : 2^\Omega \to \mathbb{R}$*. The* global submodularity distance $\mathcal{E}$ *is defined by* $\mathcal{E}[f] := \max_{\mathcal{A},\mathcal{B}\subseteq\Omega} f(\mathcal{A}\cup\mathcal{B}) + f(\mathcal{A}\cap\mathcal{B}) - f(\mathcal{A}) - f(\mathcal{B})$.

The global submodularity distance is a general purpose metric; we demonstrate its value in identifying operations that preserve approximate submodularity and proving general results about set functions (Section 2.3). The remaining metrics are inspired by a characterization of increasing, submodular functions.

Lemma 1. *(Edmonds 1970) Let $f : 2^\Omega \to \mathbb{R}$. Then f is increasing and submodular if and only if for any $\mathcal{A}, \mathcal{B} \subseteq \Omega, s \in \Omega$, $f(\mathcal{A} \cup \mathcal{B} \cup \{s\}) - f(\mathcal{A} \cup \mathcal{B}) \leq f(\mathcal{A} \cup \{s\}) - f(\mathcal{A})$.*

We note that Edmonds (1970) provides a characterization that is equivalent to Lemma 1, though the presentation is slightly different. Using Lemma 1, we present a metric, termed the *pairwise violation*, which gives global information as to how far a function is from being submodular.

Definition 5. *Let $f : 2^\Omega \to \mathbb{R}$. Consider $\ell \in \{0, ..., |\Omega| - 1\}, k \in \{0, ..., |\Omega|\}$. The (ℓ, k)-*pairwise violation *of f is defined as*

$$d^{\ell,k}[f] \coloneqq \max_{\substack{\mathcal{A},\mathcal{B}\subseteq\Omega, s\in\Omega \\ |\mathcal{A}|=\ell, |\mathcal{B}|=k}} f(\mathcal{A} \cup \mathcal{B} \cup \{s\}) - f(\mathcal{A} \cup \mathcal{B}) - f(\mathcal{A} \cup \{s\}) + f(\mathcal{A}).$$

Thus, the pairwise violation represents the worst-case violation of the condition in Lemma 1 given $\mathcal{A}$ and $\mathcal{B}$ with fixed cardinalities. We return to the sensor example from Section 2.1 to provide intuition for Definition 5. The (ℓ, k)-pairwise violation captures the case in which a single sensor added to a sparse sensor network (given by $\mathcal{A}$) creates a smaller marginal increase in information than when the same sensor is added to a denser network ($\mathcal{A} \cup \mathcal{B}$). Note that $d^{\ell,k}[f] \leq 0$ for all ℓ and k if and only if f is increasing and submodular. Also, for any f, $d^{\ell,k}[f] \leq \mathcal{E}[f]$.

We present two metrics using pairwise violations; the first is the maximum of all pairwise violations and is useful in general approximately submodular analysis, and the second considers sums of pairwise violations, which we use for greedy algorithm analysis.

Definition 6. *Let $f : 2^\Omega \to \mathbb{R}$. The* marginal violation *of f is defined as $D[f] \coloneqq \max\{d^{\ell,k}[f] \mid \ell \in \{0, \ldots, |\Omega| - 1\}, k \in \{0, \ldots, |\Omega|\}\}$.*

Note that $d^{0,0}[f] = 0$, so $D[f] \geq 0$.

Definition 7. *Let $f : 2^\Omega \to \mathbb{R}$, $L \in \{0, ..., |\Omega| - 1\}, K \in \{1, ..., |\Omega|\}$ and $\delta^{\ell,K}[f] \coloneqq \sum_{k=0}^{K-1} d^{\ell,k}[f]$, for any $\ell \in \{0, \ldots, L\}$. The* (L, K)-submodularity violation *of f is*

$$\Delta^{L,K}[f] \coloneqq \max_{\ell\in\{0,\ldots,L\}} \delta^{\ell,K}[f].$$

Although an understanding of the global behavior of a function is generally useful, many algorithms only call for function evaluations within a subset of the domain, i.e., by effectively utilizing local information of the underlying function. Thus, we also provide variants of d, δ, and Δ to focus on submodularity violations when $\mathcal{A}$ is fixed to a set or is chosen from a collection of sets.

Definition 8. *For a subset $\mathcal{A} \subseteq \Omega, k \in \{0, \ldots, |\Omega|\}$, the $(\mathcal{A}, k)$-*local pairwise violation *is*

$$\hat{d}^{\mathcal{A},k}[f] \coloneqq \max_{\substack{\mathcal{B} \subseteq \Omega, |\mathcal{B}|=k, \\ s \in \Omega}} f(\mathcal{A} \cup \mathcal{B} \cup \{s\}) - f(\mathcal{A} \cup \mathcal{B}) - f(\mathcal{A} \cup \{s\}) + f(\mathcal{A}).$$

Definition 9. *Given $f : 2^\Omega \to \mathbb{R}$, let $\mathscr{C}$ be a collection of subsets of Ω, $K \in \{1, ..., |\Omega|\}$, and $\hat{\delta}^{\mathcal{A},K}[f] \coloneqq \sum_{k=0}^{K-1} \hat{d}^{\mathcal{A},k}[f]$, for any $\mathcal{A} \in \mathscr{C}$. The* local $(\mathscr{C}, K)$-submodularity violation *is*

$$\hat{\Delta}^{\mathscr{C},K}[f] \coloneqq \max_{\mathcal{A} \in \mathscr{C}} \hat{\delta}^{\mathcal{A},K}[f].$$

Remark 1. *Note that $\hat{d}^{\mathcal{A},k}[f] \leq d^{|\mathcal{A}|,k}[f]$, and $\hat{\delta}^{\mathcal{A},K}[f] \leq \delta^{|\mathcal{A}|,K}[f]$, for all $\mathcal{A}, \mathcal{S} \subseteq \Omega, k \in \{0, ..., |\Omega| - 1\}, K \in \{1, ..., |\Omega|\}$. Moreover, if $\max_{\mathcal{A} \in \mathscr{C}} |\mathcal{A}| = L$, then $\hat{\Delta}^{\mathscr{C},K}[f] \leq \Delta^{L,K}[f]$.*

Further, by defining $\Omega_\mathcal{S} \coloneqq \{\mathcal{A} \subseteq \Omega \mid 1 \geq |\mathcal{A} \backslash \mathcal{S}|\}$, for any $\mathcal{S} \subseteq \Omega$, we can generalize Horel and Singer (2016)'s notion of approximate submodularity.

Definition 10. *Given $\mathcal{S} \subseteq \Omega$, a set function $f : 2^\Omega \to \mathbb{R}$ is* $(\mathcal{S}, \epsilon)$-approximately submodular *if f is ϵ-approximately submodular over $\Omega_\mathcal{S}$.*

Recall that Zhou and Spanos (2016) derive a bound by considering the local submodularity index for all location sets $\mathcal{S}$ that are subsets of the output of the greedy algorithm. However, we can further localize the authors' approximate submodularity metric, which leads to a reduction of the arguments considered for the local submodularity index.

Definition 11. *For a set function $f : 2^\Omega \to \mathbb{R}$, the* (localized) submodularity indicator *for a collection $\mathscr{C}$ of subsets of Ω with maximum cardinality K is*

$$\widehat{\mathcal{I}}^{\mathscr{C},K}[f] \coloneqq \max_{\substack{\mathcal{A} \in \mathscr{C}, \mathcal{B} \subseteq \Omega, \\ \mathcal{B} \cap \mathcal{A} = \emptyset, 2 \leq |\mathcal{B}| \leq K}} \phi^{\mathcal{A},\mathcal{B}}[f]. \tag{2}$$

If in (2) *no such $(\mathcal{A}, \mathcal{B})$ exist, then $\widehat{\mathcal{I}}^{\mathscr{C},K}[f] = 0$.*

The justification for the last sentence of Definition 11 is similar to that of Definition 2. In Section 5, we use the submodularity indicator and other metrics to provide bounds for the greedy algorithm.

There are important trade-offs between all of the metrics in this section. For instance, the more global metrics such as $D[f]$ can give insight into how close a function is to having a useful property (submodularity) and may be of use with other applications and analyses outside of the greedy algorithm (see Sections 2.3, 3, 4, and 6). On the other hand, the local metrics $\hat{\Delta}^{\mathscr{C}_L,K}[f]$ and $\widehat{\mathcal{I}}^{\mathscr{C},K}[f]$ provide the necessary information to derive bounds for the greedy algorithm (see Sections 5 and 6), yet they may be less applicable in other uses because of their local nature.

2.3 Preserving Approximate Submodularity

We prove some properties of our proposed approximate submodularity metrics (Theorem 1), as well as operations from which bounds or exact values of approximate submodularity metrics can be inferred immediately (Proposition 1). The former compares properties of our approximate submodularity metrics to true metrics. The latter concept can be thought of as "approximate submodularity preservation". Some of these results have analogs for submodular functions (for reference, see Nemhauser and Wolsey (1988), Narayanan (1997), and Bach (2013)), but others are specific to approximate submodularity. We let $\mathcal{F}$ (resp., $\mathcal{F}_+$) be the set functions (resp., that are nonnegative and increasing) over ground set Ω.

Theorem 1. *Consider a nonnegative, increasing set function $f : 2^\Omega \to \mathbb{R}$ and a metric of approximate submodularity $\zeta : \mathcal{F}_+ \to \mathbb{R}$ where ζ is defined by any of the following: (I) $\zeta[f] = \mathcal{E}[f]$, (II) $\zeta[f] = D[f]$, (III) $\zeta[f] = d^{\ell,k}[f]$, for some $\ell \in \{0, \ldots, |\Omega| - 1\}, k \in \{0, \ldots, |\Omega|\}$, or (IV) $\zeta[f] = \Delta^{L,K}[f]$, for some $L \in \{0, ..., |\Omega| - 1\}, K \in \{1, ..., |\Omega|\}$. Then we have:*

(i) The function ζ is sublinear. That is, ζ is subadditive (i.e., $\zeta[f_1] + \zeta[f_2] \geq \zeta[f_1 + f_2]$) and positively homogeneous with degree 1 (i.e., $\alpha\zeta[f] = \zeta[\alpha f]$, for $\alpha \in \mathbb{R}_+$).

(ii) For cases (I), (II), and (III), if f is not submodular, then for any $\epsilon \in [0, \zeta[f])$, there does not exist a nonnegative, increasing, submodular function $g : 2^\Omega \to \mathbb{R}$ such that $||g - f||_\infty < \frac{\epsilon}{4}$.

The contrapositive of Claim (ii) of Theorem 1 can be read as a necessary condition, which can, in some cases, remove the need for testing whether any function near f is submodular (e.g, Seshadri and Vondrák (2014)). Although our notions of approximate submodularity are not "metrics" in the analytical sense, Theorem 1 proves that they are sublinear. Sublinear functions are well studied in the literature and are the "next simplest convex functions" after affine functions (Hiriart-Urruty and Lemaréchal 2001). By definition, all metrics are sublinear. Subadditivity and positive homogeneity independently have multiple implications. They can be used to verify that a function $f_1 + f_2$ (or αf, for $\alpha \in \mathbb{R}_+$) satisfies conditions in hypotheses of results in Sections 3-5. We remark that subadditivity (and hence, sublinearity) is not a trivial property of approximate submodularity metrics in the literature; Example 2 shows the submodularity ratio is not subadditive.

We can also relate our metrics to functions, such as asymmetric seminorms, that share more properties with analytical metrics.

Definition 12. *(Cobzaş 2013) A function $\zeta : \mathbb{R}^n \to \mathbb{R}$ is an asymmetric seminorm if it is nonnegative, positively homogeneous, and subadditive.*

Corollary 1. *Let $\mathcal{E}_+ : \mathcal{F} \to \mathbb{R}$ be defined by $\mathcal{E}_+[f] \coloneqq \max\{0, \mathcal{E}[f]\}$. Then $\mathcal{E}_+$ is an asymmetric seminorm on $\mathcal{F}$.*

We provide some examples in which functions induced by an approximately submodular function inherit approximate submodularity. Denote the complement of $\mathcal{S} \subseteq \Omega$ by $\mathcal{S}^c$. Given a set function f, define $f_1, f_2 : 2^\Omega \to \mathbb{R}$ by $f_1(\mathcal{S}) = f(\mathcal{S}^c), f_2(\mathcal{S}) = f(\mathcal{S}) + f(\mathcal{S}^c) - f(\Omega)$; thus, f_2 is a symmetric, nonnegative function. Given $\mathcal{A} \subseteq \Omega$, define $f_\mathcal{A} : 2^{\Omega \setminus \mathcal{A}} \to \mathbb{R}$ by $f_\mathcal{A}(\mathcal{S}) = f(\mathcal{A} \cup \mathcal{S})$. Given a factor q of $|\Omega|$, let $\Omega(q) = \{1, \ldots, \frac{|\Omega|}{q}\}$, $\mathcal{S}(i) = \{(i-1)q + 1, \ldots, iq\}$, for all $i \in \Omega(q)$, and $f_q : 2^{\Omega(q)} \to \mathbb{R}$ be defined by $f_q(\mathcal{S}) = f(\bigcup_{i \in \mathcal{S}} \mathcal{S}(i))$. Finally, let $g : 2^\Omega \to \mathbb{R}$ be a *modular* function (g and $-g$ are submodular), and define the convolution of f and g as $f \circledast g(\mathcal{S}) = \min_{\mathcal{Z} \subseteq \mathcal{S}} f(\mathcal{Z}) + g(\mathcal{S} \setminus \mathcal{Z})$. Note $f \circledast g(\mathcal{S}) = g \circledast f(\mathcal{S})$.

Proposition 1. *Given $f : 2^\Omega \to \mathbb{R}$ and the corresponding functions $f_1, f_2, f_\mathcal{A}$, and f_q, we have: (i) $\mathcal{E}[f] = \mathcal{E}[f_1]$. (ii) $2\mathcal{E}[f] \geq \mathcal{E}[f_2]$. (iii) $\mathcal{E}[f] \geq \mathcal{E}[f_\mathcal{A}]$ (iv) $\mathcal{E}[f] \geq \mathcal{E}[f_q]$. (v) $\mathcal{E}[f] \geq \mathcal{E}[f \circledast g]$.*

Proposition 1 can be used in a fashion similar to Theorem 1. In addition, Proposition 1.(iv) can provide guarantees on a greedy algorithm that selects among prescribed subsets of elements. Our proof for Proposition 1.(v) follows similar arguments to that of Narayanan (1997), who established this result in the case of a submodular function. We note that we slightly abuse notation in Proposition 1.(iii)-(iv) as the domains of $f_\mathcal{A}$ and f_q are not 2^Ω.

3 Extensions of Approximately Submodular Functions

Next, we study extensions of approximately submodular functions that are increasing and normalized ($f(\emptyset) = 0$). Given a set function $f : 2^\Omega \to \mathbb{R}$, an *extension* of f over $[0,1]^{|\Omega|}$ is a function $F : [0,1]^{|\Omega|} \to \mathbb{R}$ such that $F(x(\mathcal{S})) = f(\mathcal{S})$, for all $\mathcal{S} \subseteq \Omega$, where $x(\mathcal{S})$ is the characteristic vector of $\mathcal{S}$. Our main results in this section are: (1) the Lovász extension is near the convex closure (Theorem 2), and (2) the multilinear extension is approximately up-concave (Theorem 3) when the underlying set function is approximately submodular. We observe that a main component of multiple key results in this section is the marginal violation D (Definition 6), which is a general and global metric of approximate submodularity. Some other results on extensions of set functions are presented in Lovász (1983), Murota (1998), and Iyer and Bilmes (2015), among others.

The Lovász (1983) extension of a set function $f : 2^{\Omega} \to \mathbb{R}$ is defined by $F^L : [0,1]^{|\Omega|} \to \mathbb{R}$ such that $F^L(x) \coloneqq \sum_{k=0}^{|\Omega|} \lambda_k f(\mathcal{C}_k)$, where $\emptyset = \mathcal{C}_0 \subset \mathcal{C}_1 \subset \cdots \subset \mathcal{C}_{|\Omega|} = \Omega$ is a chain such that $\sum_{k=0}^{|\Omega|} \lambda_k x(\mathcal{C}_k) = x$, with $\sum_{k=0}^{|\Omega|} \lambda_k = 1$, and $\lambda \geq 0$. It is well known that by defining a permutation $(\pi_1, \ldots, \pi_{|\Omega|})$ such that $x_{\pi_1} \geq x_{\pi_2} \cdots \geq x_{\pi_{|\Omega|}}$, $\mathcal{C}_0 = \emptyset, \mathcal{C}_k = \mathcal{C}_{k-1} \cup \{\pi_k\}$, for $k \in \{1, \ldots, |\Omega|\}$, the Lovász extension is equivalently defined as $F^L(x) = \sum_{k=1}^{|\Omega|} x_{\pi_k}(f(x(\mathcal{C}_k)) - f(x(\mathcal{C}_{k-1})))$ (see Bach (2013)).

The convex closure of f is the unique convex function $F^C : [0,1]^{|\Omega|} \to \mathbb{R}$ such that $F^C(x(\mathcal{S})) \leq f(\mathcal{S})$ for all $\mathcal{S} \subseteq \Omega$ and $F^C(x) \geq G(x)$ for any other convex understimator $G : [0,1]^{|\Omega|} \to \mathbb{R}$ of f. Lovász (1983) shows that f is submodular if and only if F^L is convex; in fact, in this special case, the convex closure and the Lovász extension are equal ($F^C = F^L$). This property is useful in that the convex closure is generally difficult to compute in comparison to the Lovász extension. Although the Lovász extension does not equal the convex closure when f is not submodular, we prove a generalized result when f is approximately submodular.

Consider the following linear program parametrized by $x \in [0,1]^{|\Omega|}$:

$$V(x) = \min_y \left\{ \sum_{\mathcal{S} \subseteq \Omega} f(\mathcal{S}) y(\mathcal{S}) \;\middle|\; \sum_{\mathcal{S} \ni s} y(\mathcal{S}) = x_s, \forall\, s \in \Omega, \sum_{\mathcal{S} \subseteq \Omega} y(\mathcal{S}) = 1, y \geq 0 \right\}. \tag{3}$$

Proposition 2. *(Bach 2013) For any set function $f : 2^{\Omega} \to \mathbb{R}$ such that $f(\emptyset) = 0$, we have $V(x) = F^C(x)$, for all $x \in [0,1]^{|\Omega|}$.*

The proof of Proposition 2 relies on the Fenchel bi-conjugate of $\tilde{f} : [0,1]^{|\Omega|} \to \mathbb{R}$, where $\tilde{f}(x) = f(\mathcal{S})$ if $x = x(\mathcal{S})$, for $\mathcal{S} \subseteq \Omega$ and $+\infty$ otherwise.

Given a permutation $(\pi_1, \cdots, \pi_{|\Omega|})$, define $\mathcal{S}_0^{\pi} \coloneqq \emptyset \subset \mathcal{S}_1^{\pi} \coloneqq \{\pi_1\} \cdots \mathcal{S}_k^{\pi} \coloneqq \{\pi_1, \ldots, \pi_k\} \cdots \subset \mathcal{S}_{|\Omega|}^{\pi} \coloneqq \Omega$. Define the set $\Gamma(f)$ by

$$\Gamma(f) \coloneqq \{\gamma \in \mathbb{R}^{|\Omega|} \mid \exists \text{ permutation } \pi \text{ such that } \gamma_{\pi_i} = f(\mathcal{S}^{\pi_i}) - f(\mathcal{S}^{\pi_{i-1}}), \forall\, i \in \Omega\}.$$

Definition 13. *A function $F : [0,1]^{|\Omega|} \to \mathbb{R}$ is* ϵ-approximately convex, *if $F(\lambda x + (1-\lambda) y) \leq \epsilon + \lambda F(x) + (1-\lambda) F(y)$, for any $\lambda \in [0,1]$.*

Theorem 2. *For any increasing set function $f : 2^{\Omega} \to \mathbb{R}$ such that $f(\emptyset) = 0$,*

$$\begin{aligned} F^L(x) = \max_{\gamma \in \Gamma(f)} \sum_{s \in \Omega} \gamma_s x_s &\leq F^C(x) + |\Omega| D[f] \\ &\leq F^L(x) + |\Omega| D[f] = \max_{\gamma \in \Gamma(f)} \sum_{s \in \Omega} \gamma_s x_s + |\Omega| D[f]. \end{aligned}$$

Hence, $F^L(x) \geq F^C(x) \geq F^L(x) - |\Omega| D[f]$, and $||F^L - F^C||_{\infty} \leq |\Omega| D[f]$. Moreover, F^L is $|\Omega| D[f]$-

approximately convex.

In addition, if for some $\epsilon > 0$, F^L is ϵ-approximately convex, then $D[f] \leq \epsilon$.

Theorem 2 states that for functions with small violations of submodularity, the Lovász extension and convex closure are close to each other (with respect to the supremum norm). More generally, the approximate submodularity of f implies the approximate convexity of F^L and vice-versa. The proof of Theorem 2 uses the well-known linear program (3), but a key difference is we construct feasible primal-dual solutions with a duality gap due to the generalization to approximate submodularity.

Next, we consider the case in which there exists a submodular function g close to f. In this case, we show that the Lovász extension of g approximates the Lovász extension of f.

Proposition 3. *Given set functions $f, g : 2^\Omega \to \mathbb{R}$, where $f(\emptyset) = g(\emptyset) = 0$, and their respective Lovász extensions $F^L, G^L : [0,1]^{|\Omega|} \to \mathbb{R}$, $||F^L - G^L||_\infty = ||f - g||_\infty$.*

Thus, the approximating function g (which may be submodular) can lead to approximation methods in the discrete domain or over the hypercube using convex optimization methods.

The multilinear extension can also be used to extend a set function to $[0,1]^{|\Omega|}$ (e.g., (Chekuri et al. 2014)). The multilinear extension of f is defined by $F^M(x) := \sum_{\mathcal{S} \subseteq \Omega} f(\mathcal{S}) \prod_{i \in \mathcal{S}} x_i \prod_{i \notin \mathcal{S}} (1 - x_i)$.

Definition 14. *A function $F : [0,1]^{|\Omega|} \to \mathbb{R}$ is* up-concave *if, for any $u \in \mathbb{R}_+^{|\Omega|}, x \in [0,1]^{|\Omega|}$, $G_{x,u} : [0,1] \to \mathbb{R}$ defined by $G_{x,u}(t) = F(tu + x)$ satisfies*

$$G_{x,u}(\lambda t_1 + (1-\lambda)t_2) \geq \lambda G_{x,u}(t_1) + (1-\lambda) G_{x,u}(t_2), \tag{4}$$

where $t_1, t_2 \in \mathbb{R}$, $t_1 u + x, t_2 u + x$, and $x + \lambda t_1 u + (1-\lambda) t_2 u$ are in $[0,1]^{|\Omega|}$. For any $\epsilon \geq 0$, F is ϵ-up-concave *if (4) is replaced with*

$$G_{x,u}(\lambda t_1 + (1-\lambda)t_2) \geq \lambda G_{x,u}(t_1) + (1-\lambda) G_{x,u}(t_2) - \epsilon. \tag{5}$$

When the set function f is submodular, its multilinear extension F^M is up-concave (Chekuri et al. 2014). This has led to its use in submodular maximization in various contexts, including analysis of continuous greedy approaches and approximability results (Feldman et al. 2011, Vondrák 2013, Chekuri et al. 2014, Ene and Nguyen 2016).

The following consider the multilinear extension for approximately submodular set functions. This analysis relies on some general results we prove about ϵ-up-concavity (Propositions 4 and 5).

Proposition 4. *Let $F : [0,1]^n \to \mathbb{R}$ be a differentiable function such that for any $x, y \in [0,1]^n$ such that $x - y \in \mathbb{R}_+^n \cup \mathbb{R}_-^n$, we have $F(y) \leq F(x) + \nabla F(x)^\top (y - x) + \epsilon$, where $\epsilon > 0$. Then F is ϵ-up-concave.*

Proposition 5. *Let $F : [0,1]^n \to \mathbb{R}$ be a twice-differentiable function with $\nabla^2 F(x)_{i,i} = 0$ and $\nabla^2 F(x)_{i,j} \leq \omega \in \mathbb{R}_+$. Then, F is $(\frac{n}{2}(n^{3/2}-1)\omega)$-up-concave.*

Theorem 3. *The multilinear extension F^M of f is $(|\Omega|(|\Omega|^{3/2}-1)2^{|\Omega|-4}D[f])$-up-concave.*

The first step of the proof of Theorem 3 uses a second-order condition similar to that of Vondrak (2008), but it also includes multiple additional steps to address the approximate submodularity.

4 Valid Inequalities of Polyhedra Associated With Approximately Submodular Functions

We use approximate submodularity metrics from Section 2.2 to derive valid inequalities for some mixed-integer sets. Our analyses are similar to analogs in submodular analysis (Atamtürk and Narayanan (2008), Atamtürk and Narayanan (2009), and Atamtürk and Narayanan (2020)) with additional details to generalize to approximate submodularity.

4.1 Epigraph Inequalities

We study the epigraphs of set functions. These mixed-integer sets can be useful when minimizing a submodular function (Atamtürk and Gómez 2020). We consider the case when the function is approximately submodular. Let $\phi : \mathbb{R} \to \mathbb{R}_+$ be increasing, and for any $\tau \in \mathbb{R}_+$, let $F_\tau : [0,1]^{|\Omega|} \to \mathbb{R}_+$ be defined by $F_\tau(x) = \phi(\tau + \sum_{i\in\Omega} c_i x_i)$, where $c \in \mathbb{R}_+^{|\Omega|}$. Thus, F_τ is increasing. Consider the mixed-integer feasible region $H_{\mathbb{B}} = \left\{(x,z) \in \mathbb{B}^{|\Omega|} \times \mathbb{R}_+ \mid F_\sigma(x) \leq z\right\}$, where $\sigma \geq 0$, and $c \in \mathbb{R}_+^{|\Omega|}$. Define the set function $f_\tau : 2^\Omega \to \mathbb{R}$ by $f_\tau(\mathcal{S}) = F_\tau(x(\mathcal{S}))$. Notice that f_τ is increasing and $f_\tau(\emptyset) = 0$ if and only if $\phi(\tau) = 0$. Therefore, define $g_\tau : 2^\Omega \to \mathbb{R}$ by $g_\tau(\mathcal{S}) = f_\tau(\mathcal{S}) - \phi(\tau)$, which is normalized, $g_\tau(\emptyset) = 0$, and is increasing; hence, it is also nonnegative. Note that $D[f_\tau] = D[g_\tau]$. We denote the Lovász extension of g_τ by G_τ^L.

For any $\gamma \in \Gamma(g_\tau)$—i.e., $\gamma_{\pi_k} = f_\tau(\mathcal{S}_k^\pi) - f_\tau(\mathcal{S}_{k-1}^\pi) = \phi(\tau + \sum_{i=1}^k c_{\pi_k}) - \phi(\tau + \sum_{i=1}^k c_{\pi_{k-1}})$ for some permutation π of $(1,\dots,|\Omega|)$—consider the inequality

$$\sum_{s\in\Omega} \gamma_s x_s \leq z - \phi(\tau). \tag{6}$$

When ϕ is the square root function, then f_τ is a submodular set function, and Atamtürk and Narayanan (2008) show that inequality (6) is valid for the convex hull $\mathrm{conv}(H_{\mathbb{B}})$ for $\tau = \sigma$. In fact, along with the variable bounds, such inequalities describe $\mathrm{conv}(H_{\mathbb{B}})$. In the more general case, where ϕ is such that f_τ is approximately submodular, we show that similar inequalities are still valid for $\mathrm{conv}(H_{\mathbb{B}})$.

Lemma 2. *For any* $\gamma \in \Gamma(g_\tau), \mathcal{S} \subseteq \Omega$, *we have* $-|\Omega|D[f_\tau] + \sum_{s\in\Omega} \gamma_s(x(\mathcal{S}))_s \leq f_\tau(\mathcal{S}) - \phi(\tau)$.

Proposition 6. *For any* $\gamma \in \Gamma(g_\sigma)$, *the following inequality is valid for* $\mathrm{conv}(H_{\mathbb{B}})$*:*

$$-|\Omega|D[f_\sigma] + \sum_{s\in\Omega} \gamma_s x_s \leq z - \phi(\sigma). \tag{7}$$

Proposition 6 illustrates what is lost between submodularity and approximate submodularity in deriving valid inequalities in this setting. When f_σ and g_σ are approximately submodular, $D[f_\tau] > 0$ may lead to looser valid inequalities. Our proof of Proposition 6 follows arguments similar to those of Atamtürk and Narayanan (2008), who establish the result when ϕ is the square root function.

Next, we consider the epigraph of a general, increasing, nonnegative, approximately submodular function f, $H_f := \mathrm{conv}(\{(x,z) \in \mathbb{R}^{|\Omega|} \times \mathbb{R} \mid f(x(\mathcal{S})) \leq z\})$. Consider the associated polyhedron $P_f := \{\gamma \in \mathbb{R}^{|\Omega|} \mid \sum_{s\in\Omega} \gamma_s \leq f(\mathcal{S}), \forall\, \mathcal{S} \subseteq \Omega\}$. We refer to the variable bounds as trivial inequalities of H_f.

Proposition 7. *(Atamtürk and Narayanan 2020)*

1. *Any nontrivial facet-defining inequality* $\sum_{s\in\Omega} \gamma_s x_s \leq \alpha z + \gamma_0$ *for* H_f *satisfies* $\gamma_0 \geq 0$ *and* $\alpha = 1$ *(up to scaling).*
2. *The inequality* $\sum_{s\in\Omega} \gamma_s x_s \leq z$ *is valid for* H_f *if and only if* $\gamma \in P_f$.
3. *The inequality* $\sum_{s\in\Omega} \gamma_s x_s \leq z$ *is facet-defining for* H_f *if and only if* γ *is an extreme point of* P_f.

Atamtürk and Narayanan (2020) prove that nontrivial facets of H_f are homogeneous. We establish a similar result for approximately submodular functions.

Proposition 8. *Let* $f : 2^\Omega \to \mathbb{R}$ *be increasing with* $f(\emptyset) = 0$. *Suppose* $\gamma \in \mathbb{R}^{|\Omega|}$ *and*

$$\sum_{s\in\Omega} \gamma_s x_s \leq z + |\Omega|D[f] + \gamma_0 \tag{8}$$

defines a nontrivial facet of H_f. *Let* $\bar{f} : 2^\Omega \to \mathbb{R}$ *be defined by* $\bar{f}(\emptyset) = 0, \bar{f}(\mathcal{S}) = f(\mathcal{S}) + |\Omega|D[f] + \gamma_0$, *for all nonempty* $\mathcal{S} \subseteq \Omega$, *and suppose* $\gamma \in \Gamma(\bar{f})$. *Then,* $\gamma_0 \leq 0$.

The proof of Proposition 8 proceeds similarly to that of the submodular case in Atamtürk and Narayanan (2020), with some additional steps to account for the approximate submodularity generalization. This includes bounding $\max_{\gamma\in\Gamma[f]} \sum_{s\in\mathcal{S}} \gamma_s$ using the marginal violation D. Given the conditions in the hypothesis of Proposition 8, the constant term $|\Omega|D[f] + \gamma_0$ is bounded below by 0 and above by $|\Omega|D[f]$; when f is submodular, the condition $\gamma \in \Gamma(\bar{f})$ is implied, $D[\bar{f}] = 0$, and the nontrivial facets are homogeneous. We also remark that Atamtürk and Narayanan (2020) provide valid inequalties for general set functions, but these rely on a submodular-supermodular decomposition of f.

4.2 Knapsack Inequalities

Consider the polytope $X = \text{conv}\{x \in \mathbb{B}^n \mid f(\mathcal{S}(x)) \leq b\}$, where $\mathcal{S}(x)$ is the subset of Ω characterized by the binary vector x and f is a set function. When f is submodular, nonnegative, and increasing, X is known as the submodular knapsack polytope (Atamtürk and Narayanan 2009). A special case of the submodular knapsack polytope is the well-known linear knapsack set, and optimizing over it is NP-hard (Karp 1972). We consider the case where f is approximately submodular, nonnegative and increasing. Thus, we call the set X an *approximately submodular knapsack set*. Our focus is on deriving valid inequalities for this set. Some facets for 0-1 polytopes established by Atamtürk and Narayanan (2009) apply in our setting, we list them in Proposition 18 in the appendix.

Definition 15.

1. *The subset $\mathcal{S} \subseteq \Omega$ is a* cover *for X if $f(\mathcal{S}) > b$ and is* minimal *if $f(\mathcal{S}\backslash\{s\}) \leq b$ for all $s \in \mathcal{S}$.*

2. *Let $\pi = (\pi_1, \ldots, \pi_{|\Omega\backslash\mathcal{S}|})$ be a permutation of $\Omega\backslash\mathcal{S}$. Let $U_\pi(\mathcal{S}) = \{\pi_j \in \Omega\backslash\mathcal{S} \mid f(\mathcal{S} \cup \{\pi_1, \ldots, \pi_j) - f(\mathcal{S} \cup \{\pi_1, \ldots, \pi_{j-1}\}) \geq f(\{s\}), \forall\ s \in \mathcal{S}\}$. The* set-extension *of $\mathcal{S} \subseteq \Omega$ with respect to π is denoted by $E_\pi(\mathcal{S}) = \mathcal{S} \cup U_\pi(\mathcal{S})$.*

Proposition 9 extends the result Atamtürk and Narayanan (2009, Proposition 5) for submodular knapsack problems into the approximately submodular context.

Proposition 9. *If $\mathcal{S} \subseteq \Omega$ is a cover for X, the extended cover inequality $\sum_{s \in E_\pi(\mathcal{S})} x_s \leq |\mathcal{S}| - 1$ is valid for X if $f(\mathcal{S}) > |\Omega| D[f] + b$. In addition, the inequality defines a facet of $\{x \in X \mid x_s = 0, \forall\ s \notin E_\pi(\mathcal{S})\}$ if $\mathcal{S}$ is also a minimal cover and for each $s \in U_\pi(\mathcal{S})$, there exist $t_s, u_s \in \mathcal{S}$ such that $t_s \neq u_s$, and $f(\mathcal{S} \cup \{s\} \backslash \{t_s, u_s\}) \leq b$.*

We observe from Proposition 9 that in adapting the result for approximate submodularity, we add a condition for the extended cover inequality to be valid. The proof follows similar steps as those in Atamtürk and Narayanan (2009), except it accounts for violated submodularity inequalities.

5 Approximate Submodularity and the Greedy Algorithm

We present performance guarantees for the greedy algorithm on set function maximization subject to a cardinality constraint. Our performance guarantees are derived using the proposed metrics in Section 2.2. Let Ω be a finite set and let $f : 2^\Omega \to \mathbb{R}_+$ be an increasing function. Let $\widehat{\mathcal{S}}_K \in \arg\max_{\mathcal{S} \subseteq \Omega} \{f(\mathcal{S}) \text{ subject to } |\mathcal{S}| \leq K\}$ where K denotes the maximum cardinality, and let $\mathcal{S}_L$ be the set selected by the greedy algorithm at iteration L. We first review performance bounds in the literature (from metrics in Section 2.1), then we present new bounds based on our novel metrics.

5.1 Existing Submodularity Metrics and Bounds

The literature on approximate submodular functions has grown quickly over the last few years; we review some existing studies and refer to the works cited in Section 2.1 for additional references. Zhou and Spanos (2016) use the submodularity index (Definition 2) to derive a bound.

Proposition 10. *(Zhou and Spanos 2016) Let K be the maximum cardinality parameter and the number of iterations run by the greedy algorithm. Suppose f is a nonnegative, increasing set function, and $\mathcal{I}^{\mathcal{S}_K,K}[f] \in \Big(0, f(\widehat{\mathcal{S}}_K)\Big]$. Then*

$$f(\mathcal{S}_K) \geq \left(1 - \frac{1}{e} - \frac{\mathcal{I}(\mathcal{S}_K, K)}{f(\mathcal{S}_K)}\right) f(\widehat{\mathcal{S}}_K).$$

The authors' original result requires the size of the set as a result of the greedy algorithm to be equal to the maximum cardinality parameter, and they 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. Note that Zhou and Spanos (2016) consider submodular function maximization ($\mathcal{I}^{\mathcal{S},L}[f] \leq 0$ 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)'s submodularity ratio (Definition 3) also yields a greedy algorithm bound.

Proposition 11. *(Das and Kempe 2011) Let $K \in \{1, ..., |\Omega| - 1\}$ be the maximum cardinality parameter and L be the number of iterations run by the greedy algorithm. If f is a nonnegative, increasing set function, then*

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

Horel and Singer (2016) also consider a multiplicative bound that is global; it incorporates deviations from submodularity over the entire domain of the function by using ϵ-approximate submodularity. In Section 2.2, we generalized this notion of approximate submodularity to apply to a subset of the domain (Definition 10). We extend Horel and Singer (2016)'s bound to apply to cases when $L \neq K$ and also incorporate local approximate submodularity.

Proposition 12. *(Horel and Singer 2016) Let K be the maximum cardinality parameter and L be the number of iterations run by the greedy algorithm. Consider $\epsilon \in (0, 1)$. If f is a nonnegative, increasing, $(\mathcal{S}_L \cup \widehat{\mathcal{S}}_K, \epsilon)$-approximately submodular set function, then*

$$f(\mathcal{S}_L) \geq \frac{(1-\epsilon)^2 f(\widehat{\mathcal{S}}_K)}{4K\epsilon + (1-\epsilon)^2} \left(1 - \left(\frac{(K-1)(1-\epsilon)^2}{K(1+\epsilon)^2}\right)^L\right).$$

After accounting for the generalized approximate submodularity definition, the proof of Proposition 12 is similar to that of Horel and Singer (2016), who also prove that in the global case (ϵ-approximately submodular), there is a constant factor bound when $\epsilon \in O(1/k)$.

5.2 Proposed Bounds and Global-Local Trade-off

We propose greedy algorithm bounds for increasing, nonnegative, and approximately submodular functions. Each bound adheres to three criteria: (1) it is well-defined everywhere, (2) its result holds when the number of greedy algorithm 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 criterion (1). Criterion (2) allows for flexibility in the type of approximate solution found by the greedy algorithm. One can terminate the greedy algorithm in fewer iterations than the cardinality parameter to obtain a sparse solution. On the other hand, because the greedy algorithm is relatively inexpensive compared to exhaustive search, one can quickly obtain a dense solution with additional iterations beyond the cardinality constraint. Using the approximate submodularity metrics in Section 2.2, we derive new bounds that fit these guidelines. We emphasize that all of our new bounds (Theorems 4, 5, and 6) for the greedy algorithm apply to any increasing, nonnegative set function. 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 past research in the area (see Das and Kempe 2018, Remark 7).

Recall that the submodularity violation metric (Definition 7) can be interpreted as a sum of worst-case pairwise violations given the parameters L and K. We now derive performance bounds for the greedy algorithm at iteration L with maximum cardinality parameter K.

Theorem 4. *Let $K \in \{1, ..., |\Omega|\}$ be the maximum cardinality parameter and $L \in \{0, ..., |\Omega| - 1\}$ be the number of iterations run by the greedy algorithm. If f is a nonnegative, increasing set function, then*

$$f(\mathcal{S}_L) \geq \left[f(\widehat{\mathcal{S}}_K) - \min\{\Delta^{L,K}[f], f(\widehat{\mathcal{S}}_K)\}\right]\left[1 - \left(\frac{K-1}{K}\right)^L\right].$$

Moreover, the above bound is tight.

The uncapacitated facility location instance from Cornuéjols et al. (1977) summarized in Figure 3 is an example of a tight instance for the above bound. Theorem 4 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 only provide a bound for $L \in \{0, ..., |\Omega| - 1\}$ as trivially, $f(\mathcal{S}_{|\Omega|}) = f(\Omega)$. The min operation is only necessary for functions that are quite far from submodular; none of the numerical examples in this paper require the min operation, and $\Delta^{L,K}[f]$ alone could equivalently replace this term in these instances. If f is nonnegative, increasing, and submodular, then the submodularity violations are always nonpositive. This enables us to state the classical greedy algorithm bound from Nemhauser et al. (1978) as a corollary of Theorem 4.

Corollary 2. *(Nemhauser et al. 1978) Let $K \in \{1,...,|\Omega|\}$ be the maximum cardinality parameter and $L \in \{0,...,|\Omega|-1\}$ be the number of iterations run by the greedy algorithm. If f is a nonnegative, increasing, submodular set function, then*

$$f(\mathcal{S}_L) \geq f(\widehat{\mathcal{S}}_K)\left[1-(1-1/K)^L\right].$$

We generalize the bound in Theorem 4 by using the local version of the submodularity violation $\hat{\Delta}$ (Definition 9). Denote the collection of subsets made by the greedy algorithm at each iteration by $\mathscr{C}_L = \{\emptyset, \mathcal{S}_1, ..., \mathcal{S}_L\}$.

Theorem 5. *Let $K \in \{1,...,|\Omega|\}$ be the maximum cardinality parameter and $L \in \{0,...,|\Omega|-1\}$ be the number of iterations run by the greedy algorithm. If f is a nonnegative, increasing set function, then*

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

where $\mathscr{C}_L$ is the collection of subsets made by the greedy algorithm at each iteration. Moreover, this bound is tight.

We now derive a new bound using the submodularity indicator (Definition 11). This results in a bound that requires fewer function calls and is valid even when the size of the set produced by the greedy algorithm differs from the maximum cardinality parameter. We use Theorem 4 to show our new bound is tight.

The number of function calls to compute the submodularity indicator is (often strictly) fewer than that of Zhou and Spanos (2016)'s submodularity index. In this sense, we have further localized the requirement for approximate submodularity, which helps produce a tighter bound (Theorem 6).

Theorem 6. *Let K be the maximum cardinality parameter and L be the number of iterations run by the greedy algorithm. If f is a nonnegative, increasing set function, then*

$$f(\mathcal{S}_L) \geq \min\left\{f(\widehat{\mathcal{S}}_K), (1-(1-1/K)^L)\left[f(\widehat{\mathcal{S}}_K) - \min\{\widehat{\mathcal{I}}^{\mathscr{C}_L,K}[f], f(\widehat{\mathcal{S}}_K)\}\right]\right\}.$$

Theorem 7. *Assume the conditions of Proposition* 10 *are satisfied, and $f(\mathcal{S}_K) > 0$. Then the bound from Theorem* 6 *is at least as tight as the bound from Proposition* 10.

From Theorem 7, one observes that for functions with positive submodularity indices, Theorem 6 strengthens Proposition 10.

Proposition 13. *Let K be the maximum cardinality parameter, and $\mathcal{A} \subseteq \Omega$, with $|\mathcal{A}| = \ell$. Then $\widehat{\mathcal{I}}^{\{\mathcal{A}\},K}[f] \leq \hat{\Delta}^{\{\mathcal{A}\},K}[f]$. Further, if $\mathscr{C} = \{\mathcal{A}_1, ..., \mathcal{A}_M\}$, for some $M \in \mathbb{N}$, then $\widehat{\mathcal{I}}^{\mathscr{C},K}[f] \leq \hat{\Delta}^{\mathscr{C},K}[f]$. In addition, if $L = \max_{m \in \{1,...,M\}} |\mathcal{A}_m|$, then $\widehat{\mathcal{I}}^{\mathscr{C},K}[f] \leq \Delta^{L,K}[f]$.*

Corollary 3. *The bound in Theorem 6 is tight.*

6 Illustrative Examples

In this section we provide examples that illustrate the results in Sections 3-5. First, we consider the approximately submodular knapsack polytope. We apply the derived valid inequalities (Proposition 9) and show that they can be used to find integer solutions when optimizing a linear function over the polytope. Second, we run the greedy algorithm on instances of the cooperative uncapacitated facility location problem, a generalization of the uncapacitated facility location problem. Notably, this new problem does not have a submodular objective function. We compare our bounds (from Section 5) to those in the literature. We show analytically and numerically that the multilinear extension of the objective function of this facility location problem is approximately up-concave.

6.1 Approximately Submodular Knapsack

Let Ω be a set of elements, $u, w \in \mathbb{R}_+^{|\Omega|} \backslash \{0\}$, $p > 1$. Let $G : [0,1]^{|\Omega|} \to \mathbb{R}$ and $H : \mathbb{R} \to \mathbb{R}$, where $G(x) = w^\top x$ and $H(z) = z^p$ if $z \geq 0$ and 0 otherwise. Define $F : [0,1]^{|\Omega|} \to \mathbb{R}$ and $f : 2^\Omega \to \mathbb{R}$, where $F(x) = u^\top x + H(G(x))$ and $f(\mathcal{S}) = F(x(\mathcal{S}))$. Because $p > 1$, H is convex and increasing. Also, G is a linear (convex) function, which implies F is convex and f is supermodular. One can observe that f is not submodular because it is not modular, and that f is increasing and nonnegative.

Proposition 14. *We have $D[f] \leq p||w||_1^{p-1}||w||_\infty$.*

Proposition 14 provides an analytical bound on $D[f]$ that can remove the need to compute $D[f]$ directly, which is helpful to verify whether Proposition 9 applies to inequalities of the form $f(\mathcal{S}) \geq |\Omega|D[f] + b$, for some $\mathcal{S} \subseteq \Omega$.

We provide an example instance in which we optimize a linear function over the integer hull of the approximately submodular knapsack polytope. Let $\Omega = \{1, 2, \ldots, 6\}$, $u = [9, 9, 9, 9, 8.85, 0]^\top$, $w = [0, 0, 0, 0, 1, 1]^\top$, $p = 1.1$, and define f, F, G, and H as stated above. Also let $c = [3, 3, 3, 3, 2, 2]^\top, b = 28.3$. Define the following instance of an approximately submodular knapsack problem (ASK) written as a binary program:

$$z_{\text{ASK}} = \max\{\sum_{s\in\Omega} c_s x_s \mid F(x) \leq b, x \in \mathbb{B}^{|\Omega|}\}. \tag{ASK}$$

The continuous relaxation of (ASK) was solved using Gurobi 9.0.1 (Gurobi Optimization LLC 2020) using a piecewise approximation of the nonlinear function with maximum absolute error of .001, with an optimal solution of $[0.0\overline{33}, 1, 1, 1, 0, 1]^\top$ and objective value 11.1. Observe that $\mathcal{S} = \{1, 2, 3, 4\}$ is a (minimal) cover. Consider the permutation of $(5, 6)$ $(\pi_1 = 5, \pi_2 = 6)$. Then $f(\mathcal{S}\cup\{\pi_1\}) - f(\mathcal{S}) = 9.85 > 9 = f(\{s\})$, and $f(\mathcal{S}\cup\{\pi_1, \pi_2\}) - f(\mathcal{S}\cup\{\pi_1\}) \approx 1.14 < f(\{s\})$, for all $s \in \mathcal{S}$. Hence, $U_\pi(\mathcal{S}) = \{5\}$ and $E_\pi(\mathcal{S}) = \{1, 2, \ldots, 5\}$. By Proposition 14, $D[f] \leq p||w||_1^{p-1}||w||_\infty \approx 1.18$, which implies $f(\mathcal{S}) = 36 > 1.18|\Omega| + 28.3 = 7.08 + 28.3 \geq |\Omega|D[f] + b$; thus, Proposition 9 implies $\sum_{s=1}^{5} x_s \leq 3$ is a valid inequality for (ASK). Solving the relaxation of (ASK) with this valid inequality yields an optimal solution of $[1, 1, 1, 0, 0, 1]^\top$ with an objective value of 11. Thus, this solution is optimal for (ASK).

In general, $f(\mathcal{S}) \geq |\Omega|D[f] + b$ does not hold, so not every extended cover inequality is valid.

6.2 Uncapacitated Facility Location

We explore how the greedy algorithm bounds in Section 5 perform empirically and illustrate the multilinear extension's approximate up-concavity. In particular, we present a generalization of the well-known uncapacitated facility location problem (see Mirchandani and Francis (1990) for a detailed overview). We choose uncapacitated facility location as a demonstrative example because of its historical importance (e.g., Cornuéjols et al. (1977)).

In the generalized uncapacitated facility location problem that we consider, the objective function is not submodular in many cases. 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 ($\Omega = \{1, ..., m\}$), 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. We note that Cornuéjols et al. (1977) consider similar conditions. Let $f : 2^\Omega \to \mathbb{R}$

be the objective function of the UFLP with cardinality parameter $K \in \{1, \dots, |\Omega|\}$

$$f(\mathcal{S}) := \begin{cases} \sum_{j=1}^{n} b_j \max_{i \in \mathcal{S}} v_{ij}, & \text{if } \mathcal{S} \neq \emptyset \\ 0, & \text{if } \mathcal{S} = \emptyset. \end{cases} \qquad \text{UFLP: } \max_{\mathcal{S} \subseteq \Omega} \{f(\mathcal{S}) \text{ subject to } |\mathcal{S}| \leq K\}.$$

Here, $\mathcal{S}$ is a subset of facility locations. Under these conditions, f is nonnegative, increasing, and submodular. We consider a generalization of UFLP where the objective function is approximately submodular function. Let $\mathcal{S}^2 = \{(p, q) \in \{1, ..., m\}^2$, for any $\mathcal{S} \subseteq \Omega$. We introduce a nonnegative reward u_{pq} associated with the simultaneous selection of facilities p and q, where $(p, q) \in \Omega^2$. We assume that $u_{pp} = 0$ for all $p \in \Omega$. Define $h : 2^\Omega \to \mathbb{R}$ as the objective function of the *cooperative uncapacitated facility location problem* (CUFLP) with maximum cardinality parameter K.

$$h(\mathcal{S}) := \begin{cases} \sum_{j=1}^{n} b_j \max_{i \in \mathcal{S}} v_{ij} + \sum_{(p,q) \in \mathcal{S}^2} u_{pq}, & \text{if } \mathcal{S} \neq \emptyset \\ 0 & \text{if } \mathcal{S} = \emptyset. \end{cases} \qquad \text{CUFLP : } \max_{\mathcal{S} \subseteq \Omega} \{h(\mathcal{S}) \text{ subject to } |\mathcal{S}| \leq K\}.$$

Remark 2. *It is well known that UFLP is NP-hard (Cornuéjols et al. 1983); thus, CUFLP (which includes UFLP as a special case) is also NP-hard.*

Remark 3. *The objective function of CUFLP is not submodular in general.*

Example 3 in the appendix illustates this remark.

Let $\text{supp}(u) := \{(p, q) \in \Omega^2 \mid u_{pq} > 0\}$.

Proposition 15. *Given an instance of CUFLP, we have* $d^{\ell,k}[h] \leq |\text{supp}(u)| \max\{u_{pq} \mid (p, q) \in \Omega^2\}$ *for all* $\ell \in \{0, ..., m - 1\}, k \in \{0, ..., m\}$.

The bound on the pairwise violations in Proposition 15 can be used to provide weaker bounds than those of Theorems 4, 5, and 6. Again, the quality of these bounds depends on the objective function's deviation from submodularity (i.e., the cooperative bonuses). In addition to greedy algorithm bounds, Proposition 15 can provide bounds for functions over the hypercube $[0, 1]^{|\Omega|}$ by using results such as Theorem 2.

Corollary 4. *Let* $H^C, H^L, H^M : [0, 1]^{|\Omega|} \to \mathbb{R}$ *be the convex closure, Lovász extension, and multilinear extension of* h*, respectively. Let* $D' = |\text{supp}(u)| \max\{u_{pq} \mid (p, q) \in \Omega^2\}$*. Then* $||H^C - H^L||_\infty \leq |\Omega| D'$*. Also,* H^M *is* $((|\Omega|^{3/2} - 1)2^{|\Omega|-4} D')$*-approximately up-concave.*

To demonstrate the proposed bounds, we consider a numerical example adapted from Cornuéjols et al. (1977) in which there are seven facilities, twelve 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 $t \in \{\frac{1}{4}, \frac{1}{2}, 1, 2\}$ to generate instances with various levels of submodularity violation. We denote the resulting objective function by h^t. As t increases, $\frac{1}{t}u$ decreases and h^t approaches the submodular function f.

Figure 1 compares the optimal objective value (triangles), the objective value of the set chosen by the greedy algorithm (rectangles), and various bounds for non-submodular functions. These include two of our proposed bounds; for ease of exposition, we refer to them as the global Delta Bound (Theorem 4, crosses) and the Indicator Bound (Theorem 6, 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 Das and Kempe (2011) (S.R. Bound, circles), Zhou and Spanos (2016) (Index Bound, Xs), and Horel and Singer (2016) (Eps Bound, stars). The Eps Bound was produced using Proposition 16.

Proposition 16. *For the above instances of cooperative facility location problems, with f as the submodular function such that $(1-\epsilon_H)f(\mathcal{S}) \leq h(\mathcal{S}) \leq (1+\epsilon_H)f(\mathcal{S})$, the smallest valid ϵ_H in Proposition 12 is $u_{6,7}/f(6,7)$.*

Proposition 16 uses f from the uncapacitated facility location problem as the submodular function that approximates h; for some other submodular function g, there may be a smaller valid ϵ_H.

When t is small, $h^t(\cdot)$ is far from submodular, in a global sense, and thus $\Delta^{L,L}(h^t)$ 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 ∞), h^t approaches f and $\Delta^{L,L}[h^t]$ decreases, which indicates that the greedy algorithm can perform reasonably well.

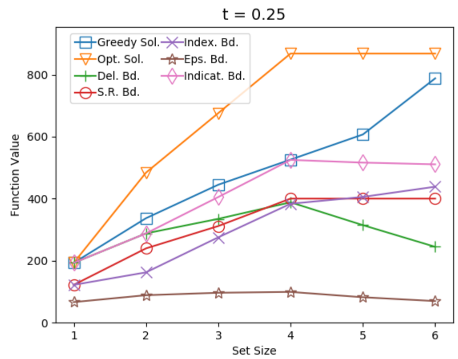
The Delta Bound and Eps Bound incorporate global information, although the former is additive while the latter is multiplicative. Still, both of these bounds are more conservative than the other local behavior bounds, generally. The Delta Bound is always above the Eps Bound in these examples; whether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

The Indicator Bound does well in comparison to other bounds in these examples. The results reinforce Theorem 7; the Indicator Bound is tighter than the Index Bound for approximately submodular optimization. Although there is a slight decrease in the Indicator Bound when the set size is large, this is also the case with the Eps Bound in some cases. The bounds are nontrivial in these examples, and they are more informative when the function is almost submodular.

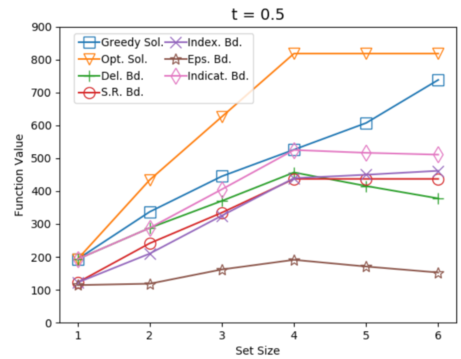
Finally, we also demonstrate the degree to which the multilinear extension of CUFLP is approximately up-concave. In Figure 2, we show the multilinear extension as a function of one or two variables (holding others fixed). We set the first two components of the input to the multilinear extension equal to 1, and all other components (except 6 and 7) to 0. Figures 2a and 2b show the multilinear extension as a function of x_6 and x_7 when $u_{6,7}$ equals 100 and 800, respectively. One can observe that when the perturbation from $u_{6,7}$ is

Figure 1: Greedy algorithm performance and bounds for CUFLP. The Delta Bound (+) is from Theorem 4, and the Indicator Bound (◊) is from Theorem 6. Notice that the relative performance of the bounds changes depending on how close the objective function is to the original submodular function.

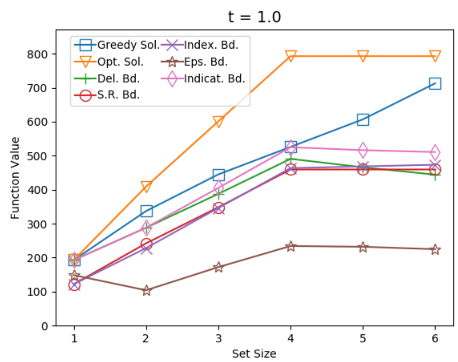
(a) As t is small, the objective is far from submodular. The (global) Delta Bound does not provide a useful guarantee, but the (local) Indicator Bound does.



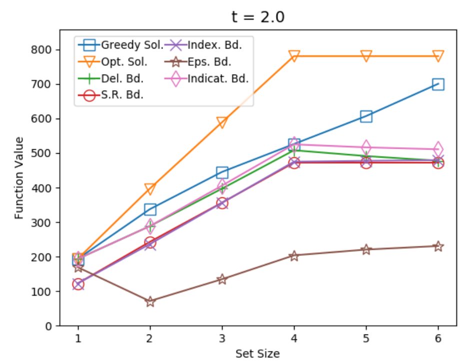
(b) As t increases, the bounds become more meaningful.



(c) The Eps Bound (Horel and Singer 2016) appears to be the most conservative when $t = 1$ for latter iterations.



(d) When $t = 2$, most of the bounds are very close together.



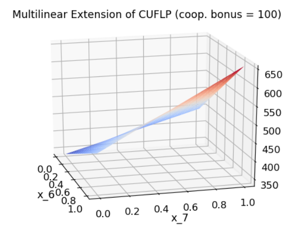
small, the multilinear extension appears nearly linear, thus almost up-concave; on the other hand, when $u_{6,7}$ is very large, the extension is further from concave. Figure 2c shows the multilinear extension from another perspective. Here, $x_6 = x_7$, and as they increase together, it is demonstrated that the multilinear extension is convex (thus, not up-concave). However, the curvature is much greater with the larger cooperative bonus, indicating the extension is closer to up-concave when the set function is closer to submodular.

7 Conclusion

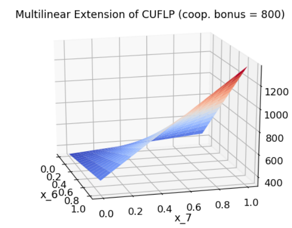
Submodularity's value in discrete optimization has long been established. Recently, notions of approximate submodularity have been applied to the greedy algorithm and similar approaches. In this work, we introduce new approximate submodularity metrics that have broad applicability in discrete optimization. We derive

Figure 2: The multilinear extension is approximately up-concave. The plots show the multilinear extension based off of the initial point $\bar{x}$ with $\bar{x}_1 = \bar{x}_2 = 1, \bar{x}_j = 0, j \in \{3,4,5,6,7\}$. In the first two plots we change the values of $\bar{x}_6$ and $\bar{x}_7$, which correspond to facility 6 and 7, between 0 and 1. With a small cooperative bonus (e.g., 4), the extension is nearly up-concave. When the bonus is large (e.g., 32), the extension is further from up-concavity. In the third plot, $\bar{x}$ is changed by $\lambda(e_6 + e_7)$.

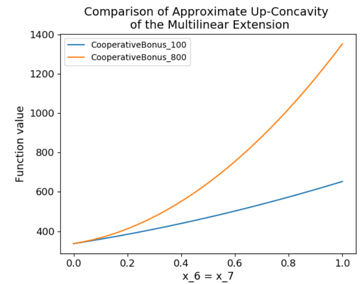
(a) When the perturbation due to the bonus is relatively small, the multilinear extension is nearly up-concave (in fact, nearly linear).



(b) For a large cooperative bonus (thus large perturbation from submodularity), the extension is further from up-concavity.



(c) The multilinear extension evaluated over $\{\hat{x} = \bar{x} + \lambda(e_6 + e_7) \mid 0 \leq \lambda \leq 1\}, \bar{x}_1 = \bar{x}_2 = 1, \bar{x}_j = 0, j \in \{3,4,5,6,7\}$.



fundamental properties about our metrics, including which set function operations preserve approximate submodularity. We establish connections between our notions of approximate submodularity and properties of well-known set function extensions, such as the approximate convexity of the Lovász extension. Our approximate submodularity metrics can directly extend analyses in areas such as greedy algorithm bounds and valid inequalities. Our illustrative examples show that we can use these valid inequalities to find integral optimal solutions to the approximately submodular knapsack problem, and we can use our metrics to produce competitive greedy algorithm bounds for a generalized facility location problem.

Acknowledgements

The authors would like to thank their colleagues, which include Seth Brown, David Mildebrath, Logan Smith, and Silviya Valeva of Rice University, for their helpful comments. This research was funded by National Science Foundation grants CMMI-1826323 and CMMI-1826297.

References

Atamtürk, A. and A. Gómez (2020). Submodularity in conic quadratic mixed 0-1 optimization. *Operations Research 68*(2), 609–630.

Atamtürk, A. and V. Narayanan (2008). Polymatroids and mean-risk minimization in discrete optimization. *Operations Research Letters 36*(5), 618 – 622.

Atamtürk, A. and V. Narayanan (2009). The submodular knapsack polytope. *Discrete Optimization 6*(4), 333 – 344.

Atamtürk, A. and V. Narayanan (2020). Submodular function minimization and polarity. arXiv 1912.13238.

Bach, F. (2013). *Learning with Submodular Functions: A Convex Optimization Perspective*. Hanover, MA, USA: Now Publishers Inc.

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

Balas, E. (1975). Facets of the knapsack polytope. *Mathematical Programming 8*, 146–164.

Bian, A. A., J. M. Buhmann, A. Krause, and S. Tschiatschek (2017). Guarantees for greedy maximization of non-submodular functions with applications. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML'17, pp. 498–507. JMLR.

Borodin, A., D. T. M. Le, and Y. Ye (2014). Weakly submodular functions. arXiv 1401.6697.

Chekuri, C., J. Vondrk, and R. Zenklusen (2014). Submodular function maximization via the multilinear relaxation and contention resolution schemes. *SIAM Journal on Computing 43*(6), 1831–1879.

Cobzaş, Ş. (2013). *Functional Analysis in Asymmetric Normed Spaces*. Basel: Birkhäuser, Basel.

Cornuéjols, G. (2008). Valid inequalities for mixed integer linear programs. *Mathematical Programming 112*, 3–44.

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

Cornuéjols, G., G. Nemhauser, and L. Wolsey (1983). The uncapacitated facility location problem. In P. E. Mirchandani and R. L. Francis (Eds.), *Discrete Location Theory*, pp. 119–171. New York, NY: John Wiley and Sons Inc.

Das, A. and D. Kempe (2011). 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, USA, pp. 1057–1064. Omnipress.

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

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

Edmonds, J. (1971, Dec). Matroids and the greedy algorithm. *Mathematical Programming 1*(1), 127–136.

Ene, A. and H. L. Nguyen (2016). Constrained submodular maximization: Beyond 1/e. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 248–257.

Feldman, M., J. Naor, and R. Schwartz (2011). A unified continuous greedy algorithm for submodular maximization. In *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science*, pp. 570–579.

Fisher, M. L., G. L. Nemhauser, and L. A. Wolsey (1978). An analysis of approximations for maximizing submodular set functions—II. In M. L. Balinski and A. J. Hoffman (Eds.), *Polyhedral Combinatorics*, pp. 73–87. Berlin, Heidelberg: Springer Berlin Heidelberg.

Fujishige, S. and K. Murota (2000). Notes on L-/M-convex functions and the separation theorems. *Mathematical Programming 88*, 129–146.

Gómez, A. (2020). Strong formulations for conic quadratic optimization with indicator variables. *Mathematical Programming (To appear)*.

Gurobi Optimization LLC (2020). Gurobi optimizer reference manual.

Hammer, P. L., E. L. Johnson, and U. N. Peled (1975). Facet of regular 0-1 polytopes. *Mathematical Programming 8*, 179–206.

Hiriart-Urruty, J.-B. and C. Lemaréchal (2001). *Fundamentals of Convex Analysis*. Berlin, Heidelberg: Springer-Verlag.

Hochbaum, D. S. (1995). A nonlinear knapsack problem. *Operations Research Letters 17*(3), 103 – 110.

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

Hubbard, J. H. and B. B. Hubbard (2009). *Vector calculus, linear algebra, and differential forms: A unified approach* (4th ed.). Ithaca, NY: Matrix Editions.

Iyer, R. and J. Bilmes (2015). Polyhedral aspects of submodularity, convexity and concavity. arXiv 1506.07329.

Karp, R. M. (1972). Reducibility among combinatorial problems. In R. Miller, J. Thatcher, and J. Bohlinger (Eds.), *Complexity of Computer Computations*, pp. 85–103. Springer US.

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

Lovász, L. (1983). Submodular functions and convexity. In *Mathematical Programming: The State of the Art*, pp. 235–257. Springer.

Mirchandani, P. B. and R. L. Francis (Eds.) (1990). *Discrete Location Theory* (1st ed.). Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley.

Murota, K. (1998). Discrete convex analysis. *Mathematical Programming 83*, 313–371.

Murota, K. and A. Shioura (2003). Quasi M-convex and L-convex functions—Quasiconvexity in discrete optimization. *Discrete Applied Mathematics 131*(2), 467–494.

Narayanan, H. (1997). *Submodular Functions and Electrical Networks*, Volume 54 of *Annals of Discrete Mathematics*. North Holland.

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

Nemhauser, G. L. and L. A. Wolsey (1978). Best algorithms for approximating the maximum of a submodular set function. *Mathematics of Operations Research 3*(3), 177–188.

Nemhauser, G. L. and L. A. Wolsey (1988). *Integer and Combinatorial Optimization*. Wiley-Interscience. Wiley.

Seshadri, C. and J. Vondrák (2014). Is submodularity testable. *Algorithmica 69*, 1–25.

Song, Y., J. R. Luedtke, and S. Küçükyavuz (2014). Chance-constrained binary packing problems. *INFORMS Journal on Computing 26*(4), 735–747.

Sviridenko, M. (2004). A note on maximizing a submodular set function subject to a knapsack constraint. *Operations Research Letters 32*(1), 41 – 43.

Van Roy, T. J. and L. A. Wolsey (1986). Valid inequalities for mixed 0-1 programs. *Discrete Applied Mathematics 14*, 199–213.

Vondrak, J. (2008). Optimal approximation for the submodular welfare problem in the value oracle model. In *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*, STOC '08, New York, NY, USA, pp. 67–74. ACM.

Vondrák, J. (2013). Symmetry and approximability of submodular maximization problems. *SIAM Journal on Computing 42*(1), 265–304.

Whitney, H. (1935). On the abstract properties of linear dependence. *American Journal of Mathematics 57*(3), 509–533.

Zhou, Y. and C. J. Spanos (2016). Causal meets submodular: Subset selection with directed information. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett (Eds.), *Advances in Neural Information Processing Systems 29*, pp. 2649–2657. Curran Associates, Inc.

A Omitted Proofs

THEOREM 1. *Consider a nonnegative, increasing set function $f : 2^\Omega \to \mathbb{R}$ and a metric of approximate submodularity $\zeta : \mathcal{F}_+ \to \mathbb{R}$ where ζ is defined by any of the following: (I) $\zeta[f] = \mathcal{E}[f]$, (II) $\zeta[f] = D[f]$, (III) $\zeta[f] = d^{\ell,k}[f]$, for some $\ell \in \{0, \ldots, |\Omega| - 1\}, k \in \{0, \ldots, |\Omega|\}$, or (IV) $\zeta[f] = \Delta^{L,K}[f]$, for some $L \in \{0, ..., |\Omega| - 1\}, K \in \{1, ..., |\Omega|\}$. Then we have:*

(i) The function ζ is sublinear. That is, ζ is subadditive (i.e., $\zeta[f_1] + \zeta[f_2] \geq \zeta[f_1 + f_2]$) and positively homogeneous with degree 1 (i.e., $\alpha\zeta[f] = \zeta[\alpha f]$, for $\alpha \in \mathbb{R}_+$).

(ii) For cases (I), (II), and (III), if f is not submodular, then for any $\epsilon \in [0, \zeta[f])$, there does not exist a nonnegative, increasing, submodular function $g : 2^\Omega \to \mathbb{R}$ such that $||g - f||_\infty < \frac{\epsilon}{4}$.

Proof: For claim (i), we prove cases (III) and (IV), and for (ii), we prove case (III), as the others are similar. Let $f_j : 2^\Omega \to \mathbb{R}_+$ be increasing set functions for $j \in \{1, 2\}$.

Consider $\zeta(\cdot) = d^{\ell,k}(\cdot), \ell \in \{0, \ldots, |\Omega| - 1\}, k \in \{0, \ldots, |\Omega|\}$. Observe that

$$\begin{aligned}
\zeta[f_1 + f_2] &= \max_{\substack{\mathcal{A},\mathcal{B}\subseteq\Omega, s\in\Omega\\ |\mathcal{A}|=\ell, |\mathcal{B}|=k}} \sum_{j=1}^{2} f_j(\mathcal{A} \cup \mathcal{B} \cup \{s\}) - f_j(\mathcal{A} \cup \mathcal{B}) - f_j(\mathcal{A} \cup \{s\}) + f_j(\mathcal{A}) \\
&\leq \sum_{j=1}^{2} \max_{\substack{\mathcal{A},\mathcal{B}\subseteq\Omega, s\in\Omega\\ |\mathcal{A}|=\ell, |\mathcal{B}|=k}} f_j(\mathcal{A} \cup \mathcal{B} \cup \{s\}) - f_j(\mathcal{A} \cup \mathcal{B}) - f_j(\mathcal{A} \cup \{s\}) + f_j(\mathcal{A}) \\
&= \zeta[f_1] + \zeta[f_2],
\end{aligned}$$

which proves subadditivity.

Let $(\mathcal{A}^*, \mathcal{B}^*, s^*) \in \arg \max_{\substack{\mathcal{A},\mathcal{B}\subseteq\Omega, s\in\Omega\\ |\mathcal{A}|=\ell, |\mathcal{B}|=k}} f(\mathcal{A} \cup \mathcal{B} \cup \{s\}) - f(\mathcal{A} \cup \mathcal{B}) - f(\mathcal{A} \cup \{s\}) + f(\mathcal{A})$. Then, for any $\alpha \in \mathbb{R}_+$, we have $(\mathcal{A}^*, \mathcal{B}^*, s^*) \in \arg \max_{\substack{\mathcal{A},\mathcal{B}\subseteq\Omega, s\in\Omega\\ |\mathcal{A}|=\ell, |\mathcal{B}|=k}} \alpha f(\mathcal{A} \cup \mathcal{B} \cup \{s\}) - \alpha f(\mathcal{A} \cup \mathcal{B}) - \alpha f(\mathcal{A} \cup \{s\}) + \alpha f(\mathcal{A})$, which implies $\zeta(\alpha f) = d^{\ell,k}(\alpha f) = \alpha\zeta[f]$, which proves positive homogeneity.

Now suppose $\zeta[f] > \epsilon$, for some $\epsilon \geq 0$. Let $g : 2^\Omega \to \mathbb{R}$ be any nonnegative, increasing set function such that $||g - f||_\infty < \epsilon/4$. Consider $(\mathcal{A}^*, \mathcal{B}^*, s^*) \in \arg \max_{\substack{\mathcal{A},\mathcal{B}\subseteq\Omega, s\in\Omega\\ |\mathcal{A}|=\ell, |\mathcal{B}|=k}} f(\mathcal{A} \cup \mathcal{B} \cup \{s\}) - f(\mathcal{A} \cup \mathcal{B}) - f(\mathcal{A} \cup \{s\}) + f(\mathcal{A})$, we have

$$\begin{aligned}
\zeta(g) &\geq g(\mathcal{A}^* \cup \mathcal{B}^* \cup \{s^*\}) - g(\mathcal{A}^* \cup \mathcal{B}^*) - g(\mathcal{A}^* \cup \{s^*\}) + g(\mathcal{A}^*) \\
&\geq f(\mathcal{A}^* \cup \mathcal{B}^* \cup \{s^*\}) - f(\mathcal{A}^* \cup \mathcal{B}^*) - f(\mathcal{A}^* \cup \{s^*\}) + f(\mathcal{A}^*) - \epsilon \\
&= \zeta[f] - \epsilon \\
&> 0,
\end{aligned}$$

which implies g is not submodular.

We now prove claim (i) for case (IV). Let $L \in \{0, \ldots, |\Omega| - 1\}, K \in \{1, \ldots, |\Omega|\}$. Observe that by case (III), for any $\ell \in \{0, \ldots, L\}$, $\delta^{\ell,K}$ is subadditive (the sum of subadditive functions is subadditive). We have

$$\begin{aligned}
\zeta[f_1 + f_2] &= \max_{\ell \in \{0,\ldots,L\}} \delta^{\ell,K}[f_1 + f_2] \\
&\leq \max_{\ell \in \{0,\ldots,L\}} \delta^{\ell,K}[f_1] + \delta^{\ell,K}[f_2] \\
&\leq \max_{\ell \in \{0,\ldots,L\}} \delta^{\ell,K}[f_1] + \max_{\ell \in \{0,\ldots,L\}} \delta^{\ell,K}[f_2] \\
&= \zeta[f_1] + \zeta[f_2].
\end{aligned}$$

In addition, $\alpha\zeta[f] = \alpha \max_{\ell \in \{0,\ldots,L\}} \delta(\ell, K)[f] = \max_{\ell \in \{0,\ldots,L\}} \delta^{\ell,K}[\alpha f] = \zeta[\alpha f]$. □

COROLLARY 1. *Let $\mathcal{E}_+ : \mathcal{F} \to \mathbb{R}$ be defined by $\mathcal{E}_+[f] := \max\{0, \mathcal{E}[f]\}$. Then $\mathcal{E}_+$ is an asymmetric seminorm on $\mathcal{F}$.*

Proof: Let $f_1, f_2 \in \mathcal{F}$. Observe that $\mathcal{E}_+[f_1] + \mathcal{E}_+[f_2] \geq \mathcal{E}[f_1] + \mathcal{E}[f_2] \geq \mathcal{E}[f_1 + f_2]$, by Theorem 1. If $\mathcal{E}[f_1 + f_2] \geq 0$, then $\mathcal{E}_+[f_1] + \mathcal{E}_+[f_2] \geq \mathcal{E}[f_1 + f_2] = \mathcal{E}_+[f_1 + f_2]$. Otherwise, $\mathcal{E}_+[f_1 + f_2] = 0$. We have $\mathcal{E}_+[f_1] \geq 0$, $\mathcal{E}_+[f_2] \geq 0$, which imply $\mathcal{E}_+[f_1] + \mathcal{E}_+[f_2] \geq 0 = \mathcal{E}_+[f_1 + f_2]$, thus proving subadditivity.

Suppose $f \in \mathcal{F}$ is such that $\mathcal{E}[f] \geq 0$. By Theorem 1, for any $\alpha \geq 0, \alpha\mathcal{E}_+[f] = \alpha\mathcal{E}[f] = \mathcal{E}[\alpha f] = \mathcal{E}_+[\alpha f]$. If $\mathcal{E}[f] < 0$, then $\mathcal{E}_+[f] = \mathcal{E}[f] = 0$. By similar arguments to those of Theorem 1, $\alpha\mathcal{E}_+[f] = 0 = \mathcal{E}_+[\alpha f]$; hence $\mathcal{E}_+$ satisfies positive homogeneity.

Clearly, $\mathcal{E}_+$ is nonnegative, which concludes the proof. □

Lemma 3. *(Narayanan 1997) For any modular set function $g' : 2^\Omega \to \mathbb{R}$ and $\mathcal{A}, \mathcal{B}, \mathcal{S}, \mathcal{T} \subseteq \Omega$ with $\mathcal{A} \cup \mathcal{B} = \mathcal{S} \cup \mathcal{T}$ and $\mathcal{A} \cap \mathcal{B} = \mathcal{S} \cap \mathcal{T}$, we have $g'(\mathcal{A}) + g'(\mathcal{B}) = g'(\mathcal{S}) + g'(\mathcal{T})$.*

PROPOSITION 1. *Given $f : 2^\Omega \to \mathbb{R}$ and the corresponding functions $f_1, f_2, f_\mathcal{A}$, and f_q, we have: (i) $\mathcal{E}[f] = \mathcal{E}[f_1]$. (ii) $2\mathcal{E}[f] \geq \mathcal{E}[f_2]$. (iii) $\mathcal{E}[f] \geq \mathcal{E}[f_\mathcal{A}]$ (iv) $\mathcal{E}[f] \geq \mathcal{E}[f_q]$. (v) $\mathcal{E}[f] \geq \mathcal{E}[f \circledast g]$.*

Proof: (i): For any $\mathcal{A}, \mathcal{B} \subseteq \Omega$, by de Morgan's laws, $(\mathcal{A} \cup \mathcal{B})^c = \mathcal{A}^c \cap \mathcal{B}^c$ and $(\mathcal{A} \cap \mathcal{B})^c = \mathcal{A}^c \cup \mathcal{B}^c$, from which the result immediately follows.

(ii): Define $\tilde{f} : 2^\Omega \to \mathbb{R}$ where $\tilde{f}(\mathcal{S}) = f(\mathcal{S}^c) - f(\Omega)$. Because $f(\Omega)$ is a constant, by (i), $\mathcal{E}[\tilde{f}] = \mathcal{E}[f]$. Thus, by Theorem 1, $\mathcal{E}[f_2] \leq \mathcal{E}[f] + \mathcal{E}[\tilde{f}] = 2\mathcal{E}[f]$.

(iii): Let $\mathcal{S}, \mathcal{T} \subseteq \mathcal{A}^c$. Then $\mathcal{S} \cup \mathcal{A}$ and $\mathcal{T} \cup \mathcal{A}$ are subsets of Ω. Hence, $f_\mathcal{A}(\mathcal{S}) + f_\mathcal{A}(\mathcal{T}) + \mathcal{E}[f] = f(\mathcal{A} \cup \mathcal{S}) + f(\mathcal{A} \cup \mathcal{T}) + \mathcal{E}[f] \geq f((\mathcal{A} \cup \mathcal{S}) \cup (\mathcal{A} \cup \mathcal{T})) + f((\mathcal{A} \cup \mathcal{S}) \cap (\mathcal{A} \cup \mathcal{T})) = f(\mathcal{A} \cup (\mathcal{S} \cup \mathcal{T})) + f(\mathcal{A} \cup (\mathcal{S} \cap \mathcal{T})) = f_\mathcal{A}(\mathcal{S} \cup \mathcal{T}) + f_\mathcal{A}(\mathcal{S} \cap \mathcal{T})$.

(iv): Consider $\mathcal{A}, \mathcal{B} \subseteq \Omega(q)$. We have $\mathcal{E}[f] + f_q(\mathcal{A}) + f_q(\mathcal{B}) = \mathcal{E}[f] + f(\bigcup_{i \in \mathcal{A}} \mathcal{S}(i)) + f(\bigcup_{i \in \mathcal{B}} \mathcal{S}(i)) \leq f(\bigcup_{i \in \mathcal{A} \cup \mathcal{B}} \mathcal{S}(i)) + f(\bigcup_{i \in \mathcal{A} \cap \mathcal{B}} \mathcal{S}(i)) = f_q(\mathcal{A} \cup \mathcal{B}) + f_q(\mathcal{A} \cap \mathcal{B})$.

(v) This proof is similar to that of Narayanan (1997). Let $\mathcal{S}, \mathcal{T}, \mathcal{Z}_{\mathcal{S}}, \mathcal{Z}_{\mathcal{T}} \subseteq \Omega$, where $f \circledast g(\mathcal{S}) = f(\mathcal{Z}_{\mathcal{S}}) + g(\mathcal{C}\backslash\mathcal{Z}_{\mathcal{S}})$, and $f \circledast g(\mathcal{T}) = f(\mathcal{Z}_{\mathcal{T}}) + g(\mathcal{T}\backslash\mathcal{Z}_{\mathcal{T}})$. It is not hard to show that $(\mathcal{S}\backslash\mathcal{Z}_{\mathcal{S}}) \cup (\mathcal{T}\backslash\mathcal{Z}_{\mathcal{T}}) = ((\mathcal{S}\cup\mathcal{T})\backslash(\mathcal{Z}_{\mathcal{S}} \cup \mathcal{Z}_{\mathcal{T}})) \cup ((\mathcal{S}\cap\mathcal{T})\backslash((\mathcal{Z}_{\mathcal{S}} \cap \mathcal{Z}_{\mathcal{T}}))$ and $(\mathcal{S}\backslash\mathcal{Z}_{\mathcal{S}}) \cap (\mathcal{T}\backslash\mathcal{Z}_{\mathcal{T}}) = ((\mathcal{S}\cup\mathcal{T})\backslash(\mathcal{Z}_{\mathcal{S}} \cup \mathcal{Z}_{\mathcal{T}})) \cap ((\mathcal{S}\cap\mathcal{T})\backslash(\mathcal{Z}_{\mathcal{S}} \cap \mathcal{Z}_{\mathcal{T}})$.

By Lemma 3,

$$g(\mathcal{S}\backslash\mathcal{Z}_{\mathcal{S}}) + g(\mathcal{T}\backslash\mathcal{Z}_{\mathcal{T}}) = g((\mathcal{S}\cup\mathcal{T})\backslash(\mathcal{Z}_{\mathcal{S}} \cup \mathcal{Z}_{\mathcal{T}})) + g((\mathcal{S}\cap\mathcal{T})\backslash(\mathcal{Z}_{\mathcal{S}} \cap \mathcal{Z}_{\mathcal{T}})).$$

By the definition of $\mathcal{Z}_{\mathcal{S}}$ and $\mathcal{Z}_{\mathcal{T}}$,

$$\begin{aligned}
& f \circledast g(\mathcal{S}) + f \circledast g(\mathcal{T}) \\
= & f(\mathcal{Z}_{\mathcal{S}}) + f(\mathcal{Z}_{\mathcal{T}}) + g((\mathcal{S}\cup\mathcal{T})\backslash(\mathcal{Z}_{\mathcal{S}} \cup \mathcal{Z}_{\mathcal{T}})) + g((\mathcal{S}\cap\mathcal{T})\backslash(\mathcal{Z}_{\mathcal{S}} \cap \mathcal{Z}_{\mathcal{T}})) \\
\geq & -\mathcal{E}[f] + f(\mathcal{Z}_{\mathcal{S}} \cup \mathcal{Z}_{\mathcal{T}}) + f(\mathcal{Z}_{\mathcal{S}} \cap \mathcal{Z}_{\mathcal{T}}) + g((\mathcal{S}\cup\mathcal{T})\backslash(\mathcal{Z}_{\mathcal{S}} \cup \mathcal{Z}_{\mathcal{T}})) + g((\mathcal{S}\cap\mathcal{T})\backslash(\mathcal{Z}_{\mathcal{S}} \cap \mathcal{Z}_{\mathcal{T}})) \\
\geq & -\mathcal{E}[f] + \min_{\mathcal{Z}\subseteq\mathcal{S}\cup\mathcal{T}} f(\mathcal{Z}) + g((\mathcal{S}\cup\mathcal{T})\backslash\mathcal{Z}) + \min_{\mathcal{Z}\subseteq\mathcal{S}\cap\mathcal{T}} f(\mathcal{Z}) + g((\mathcal{S}\cap\mathcal{T})\backslash\mathcal{Z}) \\
= & -\mathcal{E}[f] + f \circledast g(\mathcal{S}\cup\mathcal{T}) + f \circledast g(\mathcal{S}\cap\mathcal{T}).
\end{aligned}$$

Hence, $\mathcal{E}[f] \geq \mathcal{E}[f \circledast g]$. □

Lemma 4. *(Bach 2013) Given a set function f, its Lovász extension F^L is positively homogeneous of degree 1.*

Lemma 5. *Suppose $x = w + \alpha x(\mathcal{A}) \in [0,1]^{|\Omega|}$, where $w, \alpha x(\mathcal{A}), \in [0,1]^{|\Omega|}$, $\mathcal{A} \subseteq \Omega$, $\alpha \in \mathbb{R}_+$, and $x_s \geq x_{s'}$ for any $s \in \mathcal{A}, s' \in \mathcal{A}^c$. If there exists a permutation π of Ω such that $x_{\pi_1} \geq \cdots \geq x_{\pi_{|\Omega|}}$ and $w_{\pi_1} \geq \cdots \geq w_{\pi_{|\Omega|}}$, then $F^L(x) = F^L(w) + \alpha f(\mathcal{A})$.*

Proof: Let π be the ranking permutation in the hypothesis, and note that it is also a ranking permutation of $x(\mathcal{A})$; that is, $x(\mathcal{A})_{\pi_1} \geq \cdots \geq x(\mathcal{A})_{\pi_{|\Omega|}}$. We have $F^L(x) = \sum_{k=1}^{|\Omega|} x_{\pi_k}(f(x(\mathcal{S}_k^\pi)) - f(x(\mathcal{S}_{k-1}^\pi)))$, and $F^L(w) = \sum_{k=1}^{|\Omega|} w_{\pi_k}(f(x(\mathcal{S}_k^\pi)) - f(x(\mathcal{S}_{k-1}^\pi)))$, implying that $F^L(x) - F^L(w) = \sum_{k=1}^{|\Omega|} \alpha x(\mathcal{A})_{\pi_k}(f(x(\mathcal{S})_k^\pi) - f(x(\mathcal{S})_{k-1}^\pi) = F^L(\alpha x(\mathcal{A})) = \alpha F^L(x(\mathcal{A})) = \alpha f(\mathcal{A})$, where we have used the positive homogeneity of the Lovász extension (Lemma 4). □

Lemma 6. *Let $f : 2^\Omega \to \mathbb{R}$ be increasing with $f(\emptyset) = 0$, and let $\tilde{\gamma} \in \Gamma(f)$. For any $\mathcal{S} \subseteq \Omega$, $f(\mathcal{S}) \geq -|\mathcal{S}|D[f] + \sum_{s\in\mathcal{S}} \tilde{\gamma}_s$.*

Proof: Consider a permutation $(\rho_1, \ldots, \rho_{|\Omega|})$ such that $\tilde{\gamma}_{\rho_1} \geq \cdots \geq \gamma_{\rho_{|\Omega|}}$ and set $\theta^* = -|\Omega|D[f]$. We prove by induction on $|\mathcal{S}|$ that $f(\mathcal{S}) \geq -|\mathcal{S}|D[f] + \sum_{s\in\mathcal{S}} \tilde{\gamma}_s$, for all $\mathcal{S} \subseteq \Omega$. The base case is easily confirmed as $f(\emptyset) = 0 = \sum_{s\in\emptyset} \tilde{\gamma}_s$. Assume for all $\tilde{\mathcal{S}} \subseteq \Omega$ with $|\tilde{\mathcal{S}}| \leq \alpha$, $f(\tilde{\mathcal{S}}) \geq -|\tilde{\mathcal{S}}|D[f] + \sum_{s\in\tilde{\mathcal{S}}} \tilde{\gamma}_s$, and let $|\mathcal{S}| = \alpha + 1$. Also, set $k = \max\{i \mid \rho_i \in \mathcal{S}\}$. Then $\mathcal{S} \cup \mathcal{S}^{\rho}_{k-1} = \mathcal{S}^{\rho}_{k}$ and $\mathcal{S} \cap \mathcal{S}^{\rho}_{k-1} = \mathcal{S}\backslash\{\rho_k\}$. Observe that by the definition of $D[f]$,

$$\begin{aligned} f(\mathcal{S}) &\geq f(\mathcal{S} \cup \mathcal{S}^{\rho}_{k-1}) + f(\mathcal{S} \cap \mathcal{S}^{\rho}_{k-1}) - f(\mathcal{S}^{\rho}_{k-1}) - D[f] \\ &= f(\mathcal{S}^{\rho}_{k}) - f(\mathcal{S}^{\rho}_{k-1}) - D[f] + f(\mathcal{S}\backslash\{\rho_k\}) \\ &= \tilde{\gamma}_{\rho_k} + f(\mathcal{S}\backslash\{\rho_k\}) - D[f] \\ &\geq \sum_{s\in\mathcal{S}} \tilde{\gamma}_s - |\mathcal{S}|D[f], \end{aligned}$$

where the last line uses the induction hypothesis. □

Theorem 2. *For any increasing set function $f : 2^{\Omega} \to \mathbb{R}$ such that $f(\emptyset) = 0$,*

$$\begin{aligned} F^L(x) = \max_{\gamma\in\Gamma(f)} \sum_{s\in\Omega} \gamma_s x_s &\leq F^C(x) + |\Omega|D[f] \\ &\leq F^L(x) + |\Omega|D[f] = \max_{\gamma\in\Gamma(f)} \sum_{s\in\Omega} \gamma_s x_s + |\Omega|D[f]. \end{aligned}$$

Hence, $F^L(x) \geq F^C(x) \geq F^L(x) - |\Omega|D[f]$, and $||F^L - F^C||_{\infty} \leq |\Omega|D[f]$. Moreover, F^L is $|\Omega|D[f]$-approximately convex.

In addition, if for some $\epsilon > 0$, F^L is ϵ-approximately convex, then $D[f] \leq \epsilon$.

Proof: The following proof uses a version of well-known linear programming duality arguments (e.g., Lovász (1983)). We first suppose f is not submodular (hence $D[f] > 0$).

Given $x \in [0,1]^{|\Omega|}$ there exists a permutation $(\pi_1, \ldots, \pi_{|\Omega|})$ such that $x_{\pi_1} \geq \cdots \geq x_{\pi_{|\Omega|}}$. Let $x_{\pi_0} = 1$. Consider the dual of (3):

$$\max_{\gamma,\theta}\ \theta + \sum_{s\in\Omega} x_s \gamma_s \tag{9a}$$

$$\text{subject to}\ \ \theta + \sum_{s\in\mathcal{S}} x_s \leq f(\mathcal{S}), \forall\ \mathcal{S} \subseteq \Omega. \tag{9b}$$

Define $y^* \in \mathbb{R}^{|2^{\Omega}|}$ by

$$y^*_{\mathcal{S}} = \begin{cases} x_{\pi_i} - x_{\pi_{i+1}},\ \text{if } \mathcal{S} = \mathcal{S}^{\pi}_i, i \in \{0, \ldots, |\Omega| - 1\}, \\ x_{\pi_{|\Omega|}},\ \text{if } \mathcal{S} = \Omega, \\ 0,\ \text{otherwise.} \end{cases}$$

We first show y^* is feasible for (3). Observe that $\sum_{s\in\Omega} y^*(\mathcal{S}) = \sum_{i=0}^{|\Omega|-1} (x_{\pi_i} - x_{\pi_{i+1}}) + x_{\pi_{|\Omega|}} = x_{\pi_0} = 1$.

In addition, for any $s \in \Omega, s = \pi_j$ for some $j \in \Omega$; hence, $\sum_{\mathcal{S} \ni s} y^*(\mathcal{S}) = \sum_{i=j}^{|\Omega|} y^*(\mathcal{S}_i^\pi) = x_{\pi_j} = x_s$. Moreover, it is easy to observe that y^* is nonnegative. Hence, y^* is feasible for (3), and $F^C(x) \leq \sum_{\mathcal{S} \subseteq \Omega} f(\mathcal{S}) y^*(\mathcal{S}) = x_{\pi_{|\Omega|}} f(\Omega) + \sum_{i=0}^{|\Omega|-1} (x_{\pi_i} - x_{\pi_{i+1}}) f(\mathcal{S}_i^\pi) = F^L(x)$.

Consider $\gamma^* \in \Gamma(f)$ such that $\gamma^*_{\pi_1} \geq \cdots \geq \gamma^*_{\pi_{|\Omega|}}$. Then $\sum_{s \in \Omega} \gamma^*_s x_s = \sum_{i=1}^{|\Omega|} (f(\mathcal{S}_i^\pi) - f(\mathcal{S}_{i-1}^\pi)) x_{\pi_i} = F^L(x)$. By Lemma 6, $(\gamma^*, -|\Omega| D[f])$ is feasible for (9).

Therefore,

$$\begin{aligned} F^L(x) &= \max_{\gamma \in \Gamma(f)} \sum_{s \in \Omega} \gamma_s x_s \\ &\leq F^C(x) + |\Omega| D[f] \\ &\leq F^L(x) + |\Omega| D[f] \\ &= |\Omega| D[f] + \max_{\gamma \in \Gamma(f)} \sum_{s \in \Omega} \gamma_s x_s. \end{aligned}$$

This also implies that $F^L(x) \geq F^C(x) \geq F^L(x) - |\Omega| D[f]$ and $||F^L - F^C||_\infty \leq |\Omega| D[f]$.

To show that F^L is approximately convex, consider $x, y \in [0,1]^{|\Omega|}, \lambda \in [0,1]$, then we have

$$\begin{aligned} F^L(\lambda x + (1-\lambda) y) &\leq F^C(\lambda x + (1-\lambda) y) + |\Omega| D[f] \\ &\leq \lambda F^C(x) + (1-\lambda) F^C(y) + |\Omega| D[f] \\ &\leq \lambda F^L(x) + (1-\lambda) F^L(y) + |\Omega| D[f]. \end{aligned}$$

For the last statement, suppose F^L is ϵ-approximately convex. For any $\mathcal{A}, \mathcal{B} \in \Omega$, denote the symmetric difference as $\mathcal{A} \dot{-} \mathcal{B} = (\mathcal{A} \backslash \mathcal{B}) \cup (\mathcal{B} \backslash \mathcal{A})$. Consider any $S, T \subseteq \Omega$, then we have

$$\begin{aligned} \frac{1}{2}(f(\mathcal{S}) + f(\mathcal{T})) + \epsilon &= \frac{1}{2} F^L(x(\mathcal{S})) + \frac{1}{2} F^L(x(\mathcal{T})) + \epsilon \\ &\geq F^L(\frac{1}{2}(x(\mathcal{S}) + x(\mathcal{T}))) \\ &= F^L(\frac{1}{2}(x(\mathcal{S} \cap \mathcal{T}) + x(\mathcal{S} \cap \mathcal{T}) + x(\mathcal{S} \dot{-} \mathcal{T}))) \\ &= F^L(\frac{1}{2}(x(\mathcal{S} \cap \mathcal{T}) + x(\mathcal{S} \cup \mathcal{T}))) \\ &= \frac{1}{2}(f(\mathcal{S} \cap \mathcal{T}) + f(\mathcal{S} \cup \mathcal{T})) \end{aligned}$$

where the last line uses the fact that the Lovász extension is positively homogeneous Lemma 4 and Lemma 5.

Now suppose that f is submodular. Then, $D[f] = 0$ and $F^C = F^L$ (e.g., see Lovász (1983), Bach (2013)). Using similar linear programming duality arguments (with $D[f]$ replaced with 0) proves that

$F^L(x) = \max_{\gamma \in \Gamma(f)} \sum_{s \in \Omega} \gamma_s x_s$. □

PROPOSITION 3. *Given set functions $f, g : 2^\Omega \to \mathbb{R}$, where $f(\emptyset) = g(\emptyset) = 0$, and their respective Lovász extensions $F^L, G^L : [0,1]^{|\Omega|} \to \mathbb{R}$, $||F^L - G^L||_\infty = ||f - g||_\infty$.*

Proof: Let $x \in [0,1]^{|\Omega|}$ and π a permutation of $(1, \ldots, |\Omega|)$ such that $x_{\pi_1} \geq x_{\pi_2} \geq \cdots \geq x_{\pi_\Omega}$, and set $x_{\pi_{|\Omega|+1}} = 0$. Then $F^L(x) = \sum_{i=1}^{|\Omega|} (f(\mathcal{S}_i^\pi) - f(\mathcal{S}_{i-1}^\pi)) x_{\pi_i} = -x_{\pi_{|\Omega|}} f(\Omega) + \sum_{i=1}^{|\Omega|-1} (x_{\pi_i} - x_{\pi_{i+1}}) f(\mathcal{S}^{\pi_i})$, and similarly for G^L. Hence,

$$\begin{aligned}
|F^L(x) - G^L(x)| &= \left| x_{\pi_{|\Omega|}} (f(\Omega) - g(\Omega)) + \sum_{i=1}^{|\Omega|-1} (x_{\pi_i} - x_{\pi_{i+1}})(f(\mathcal{S}_i^\pi) - g(\mathcal{S}_i^\pi)) \right| \\
&\leq |x_{\pi_{|\Omega|}} (f(\Omega) - g(\Omega))| + \sum_{i=1}^{|\Omega|-1} \left|(x_{\pi_i} - x_{\pi_{i+1}})(f(\mathcal{S}_i^\pi) - g(\mathcal{S}_i^\pi))\right| \\
&= |x_{\pi_{|\Omega|}}| \cdot |(f(\Omega) - g(\Omega))| + \sum_{i=1}^{|\Omega|-1} |(x_{\pi_i} - x_{\pi_{i+1}})| \cdot |(f(\mathcal{S}_i^\pi) - g(\mathcal{S}_i^\pi))| \\
&\leq |x_{\pi_{|\Omega|}}| \cdot ||f - g||_\infty + \sum_{i=1}^{|\Omega|-1} |(x_{\pi_i} - x_{\pi_{i+1}})| \cdot ||f - g||_\infty \\
&= ||f - g||_\infty \left(x_{\pi_{|\Omega|}} + \sum_{i=1}^{|\Omega|-1} (x_{\pi_i} - x_{\pi_{i+1}}) \right) \\
&\leq ||f - g||_\infty.
\end{aligned}$$

Hence, $||F^L - G^L||_\infty \leq ||f - g||_\infty$. Moreover, for some $\mathcal{S} \subseteq \Omega, ||f - g||_\infty = |f(\mathcal{S}) - g(\mathcal{S})| = |F^L(x(\mathcal{S})) - G^L(x(\mathcal{S}))|$, which implies $||F^L - G^L||_\infty = ||f - g||_\infty$. □

We provide an equivalent definition of ϵ-up-concave, which we use in our analysis. Let $F : [0,1]^{|\Omega|} \to \mathbb{R}$ and $\epsilon \geq 0$. Given $\epsilon \geq 0$, F satisfies the $(\triangledown)$ property with respect to ϵ if for any $x', y' \in \mathbb{R}^{|\Omega|}$ such that $x' - y' \in \mathbb{R}_+^{|\Omega|} \cup \mathbb{R}_-^{|\Omega|}$ and $\lambda \in [0,1]$,

$$F(\lambda x' + (1 - \lambda) y') \geq \lambda F(x') + (1 - \lambda) F(y') - \epsilon. \tag{$\triangledown$}$$

Lemma 7. *Let $F : [0,1]^{|\Omega|} \to \mathbb{R}$ and $\epsilon \geq 0$. Then F is ϵ-up-concave if and only if F satisfies the $(\triangledown)$ property with respect to ϵ.*

Proof: Suppose F satisfies the $(\triangledown)$ property. Let $x \in [0,1]^{|\Omega|}$, $u \in \mathbb{R}_+^{|\Omega|}$, $\lambda \in [0,1]$, and $t_1, t_2 \in \mathbb{R}$ such that $t_1 u + x, t_2 u + x$, and $x + \lambda t_1 u + (1 - \lambda) t_2 u$ are in $[0,1]^{|\Omega|}$. Let $x' = x + t_1 u, y' = x + t_2 u$; thus, $x' - y' \in \mathbb{R}_+^{|\Omega|} \cup \mathbb{R}_-^{|\Omega|}$. By the $(\triangledown)$ property,

$$G_{x,u}(\lambda t_1 + (1 - \lambda) t_2) = F(x + \lambda t_1 u + (1 - \lambda) t_2 u)$$

$$
\begin{aligned}
&= F(\lambda x' + (1-\lambda)y') \\
&\geq \lambda F(x') + (1-\lambda)F(y') - \epsilon \\
&= \lambda F(x + t_1 u) + (1-\lambda)F(x + t_2 u) - \epsilon \\
&= \lambda G_{x,u}(t_1) + (1-\lambda)G_{x,u}(t_2) - \epsilon.
\end{aligned}
$$

Hence, F is ϵ-up-concave.

On the other hand, suppose F is ϵ-up-concave. Let $x', y' \in [0,1]^{|\Omega|}$ such that $x' - y' \in \mathbb{R}_+^{|\Omega|} \cup \mathbb{R}_-^{|\Omega|}$. Without loss of generality, assume $x' - y' \in \mathbb{R}_-^{|\Omega|}$. Let $\lambda \in [0,1]$ and set $t_1 = 1, t_2 = 0$, $x = x'$, and $u = y' - x'$. Observe that $x + t_1 u = y', x + t_2 u = x'$, and $\lambda(x + t_1 u) + (1-\lambda)(x + t_2 u) = \lambda y' + (1-\lambda)x'$, all of which are in $[0,1]^{|\Omega|}$. By the ϵ-up-concavity of F, we have

$$
\begin{aligned}
F((1-\lambda)x' + \lambda y') &= G_{x,u}(\lambda t_1 + (1-\lambda)\lambda t_2) \\
&\geq \lambda G_{x,u}(t_1) + (1-\lambda)G_{x,u}(t_2) - \epsilon \\
&= \lambda F(y') + (1-\lambda)F(x') - \epsilon.
\end{aligned}
$$

Thus, F satisfies the $(\triangledown)$ property. □

Lemma 8. *Suppose $H \in \mathbb{R}^{n\times n}$ such that $H_{i,j} \leq \omega \in \mathbb{R}_+$ and $H_{ii} = 0$, for all $i, j \in \{1, \ldots, n\}$. Then, for all $u \in \mathbb{R}_+^n \cup \mathbb{R}_-^n$ with $||u||_2 = 1$, $u^\top H u \leq (n^{3/2} - 1)\omega$.*

Proof: Consider the matrix $W \in \mathbb{R}^{n\times n}$, where $W_{i,j} = \omega$ for all $i \neq j$ and $W_{ii} = 0$, for all $i \in \{1, \ldots, n\}$, and suppose $u \in \mathbb{R}_+^n$. Observe that $H_{i,j} - W_{i,j} \leq 0$, for all $i, j \in \{1, \ldots, n\}$, which implies $W - H$ is a copositive matrix; thus, $u \in \mathbb{R}_+^n$ implies that $u^\top (H - W)u \leq 0$. Also,

$$
\begin{aligned}
u^\top W u &= \omega \sum_{j=1}^{n} u_j(||u||_1 - u_j) \\
&\leq \omega \sum_{j=1}^{n} ||u||_1 - u_j^2 \\
&\leq \omega(n^{\frac{3}{2}} - 1)||u||_2 \\
&= \omega(n^{\frac{3}{2}} - 1).
\end{aligned}
$$

Thus, $u^\top H u \leq (n^{3/2} - 1)\omega$. If u is nonpositive, $-u$ is nonnegative and $(-u)^\top H(-u) = u^\top H u$, from which, the result follows. □

Proposition 4. *Let $F : [0,1]^n \to \mathbb{R}$ be a differentiable function such that for any $x, y \in [0,1]^n$ such that $x - y \in \mathbb{R}_+^n \cup \mathbb{R}_-^n$, we have $F(y) \leq F(x) + \nabla F(x)^\top (y - x) + \epsilon$, where $\epsilon > 0$. Then F is ϵ-up-concave.*

Proof: Let $z = \alpha y + (1-\alpha)x$, where $\alpha \in (0,1), x, y \in [0,1]^n$, and $x - y \in \mathbb{R}^n_+ \cup \mathbb{R}^n_-$; without loss of generality, assume $x - y \in \mathbb{R}^n_-$. Then $y - z \in \mathbb{R}^n_+$ and $x - z \in \mathbb{R}^n_-$, which along with the hypothesis imply $F(y) \le F(z) + \nabla F(z)^\top (y - z) + \epsilon$ and $F(x) \le F(z) + \nabla F(z)^\top (x - z) + \epsilon$. Thus, $\alpha F(y) + (1-\alpha)F(x) \le F(z) + \nabla F(z)^\top (z - z) + \epsilon = F(\alpha y + (1-\alpha)x) + \epsilon$. □

Proposition 5. *Let $F : [0,1]^n \to \mathbb{R}$ be a twice-differentiable function with $\nabla^2 F(x)_{i,i} = 0$ and $\nabla^2 F(x)_{i,j} \le \omega \in \mathbb{R}_+$. Then, F is $(\frac{n}{2}(n^{3/2} - 1)\omega)$-up-concave.*

Proof: Consider $x, y \in [0,1]^n$ such that $y - x \in \mathbb{R}^n_+ \cup \mathbb{R}^n_-$. We will first show that $F(y) \le F(x) + \nabla F(x)^\top (y - x) + \frac{n}{2}(n^{3/2} - 1)\omega$. If $x = y$, then $F(y) = F(x) + \nabla F(x)^\top (y - x)$. So assume $x \ne y$. From Taylor's theorem (see Hubbard and Hubbard (2009)), after dividing by $||y - x||_2^2$, we have, $\frac{1}{||y-x||_2^2} F(y) = \frac{1}{||y-x||_2^2} F(x) + \nabla F(x)^\top \left(\frac{y-x}{||y-x||_2^2}\right) + \frac{1}{2}\left(\frac{y-x}{||y-x||_2}\right)^\top \nabla^2 F(z) \left(\frac{y-x}{||y-x||_2}\right)$, for some $z = \alpha y + (1-\alpha)x, \alpha \in [0,1]$. Thus, $\frac{1}{||y-x||_2^2} F(y) \le \frac{1}{||y-x||_2^2} F(x) + \nabla F(x)^\top \left(\frac{y-x}{||y-x||_2^2}\right) + \frac{1}{2}(n^{3/2} - 1)\omega$, from the hypothesis and Lemma 8. Equivalently, $F(y) \le F(x) + \nabla F(x)^\top (y - x) + \frac{||y-x||_2^2}{2}(n^{3/2} - 1)\omega \le F(x) + \nabla F(x)^\top (y - x) + \frac{n}{2}(n^{3/2} - 1)\omega$. Applying Proposition 4 proves the claim. □

We note that in the proof of Proposition 5, we apply Taylor's theorem over the closed set $[0,1]^{|\Omega|}$, but it is a simple exercise to extend the domain of F^M to an open set containing the hypercube.

Theorem 3. *The multilinear extension F^M of f is $(|\Omega|(|\Omega|^{3/2} - 1)2^{|\Omega|-4} D[f])$-up-concave.*

Proof: Similar to Vondrak (2008), we prove our result using a second-order condition. From the definition of F^M, we have $\frac{\partial F^M}{\partial x_k}(x) = \sum_{\mathcal{S} \subseteq \Omega : k \in \mathcal{S}^c} (f(\mathcal{S} \cup \{k\}) - f(\mathcal{S})) \prod_{i \in \mathcal{S}} x_i \prod_{i \in \mathcal{S}^c \setminus \{k\}} (1 - x_i)$. For $\ell \ne k$, $\frac{\partial^2 F^M}{\partial x_\ell \partial x_k}(x) = \sum_{\mathcal{S} \subseteq \Omega : k, \ell \in \mathcal{S}^c} (f(\mathcal{S} \cup \{k, \ell\}) - f(\mathcal{S} \cup \{k\}) - f(\mathcal{S} \cup \{\ell\}) + f(\mathcal{S})) \prod_{i \in \mathcal{S}} x_i \prod_{i \in \mathcal{S}^c \setminus \{k, \ell\}} (1 - x_i)$.

For each $\mathcal{S} \subseteq \Omega \setminus \{k, \ell\}$, let $\mathcal{D}(\mathcal{S}) = \mathcal{S}^c \setminus \{k, \ell\}$, and notice that $\mathcal{D}(\mathcal{D}(\mathcal{S})) = \mathcal{S}$ and $\mathcal{D}(\mathcal{S}) \subseteq \Omega \setminus \{k, \ell\}$. Thus, after applying the bound $(f(\mathcal{S} \cup \{k, \ell\}) - f(\mathcal{S} \cup \{k\}) - f(\mathcal{S} \cup \{\ell\}) + f(\mathcal{S})) \le D[f]$, we can double count the terms in the second partials as

$$2\frac{\partial^2 F^M}{\partial x_\ell \partial x_k}(x) \le D[f] \sum_{\mathcal{S} \subseteq \Omega : k, \ell \in \mathcal{S}^c} \prod_{i \in \mathcal{S}} x_i \prod_{i \in \mathcal{D}(\mathcal{S})} (1 - x_i) + \prod_{i \in \mathcal{D}(\mathcal{S})} x_i \prod_{i \in \mathcal{S}} (1 - x_i).$$

If $\Omega \setminus \{k, \ell\} = \emptyset$, then the sum is equal to 0, so suppose $|\Omega \setminus \{k, \ell\}| > 1$. Consider $\mathcal{S} \subseteq \Omega \setminus \{k, \ell\}$ such that $\mathcal{S} \ne \emptyset$; thus, there exists $j \in \mathcal{S}$. Observe that

$$\prod_{i \in \mathcal{S}} x_i \prod_{i \in \mathcal{D}(\mathcal{S})} (1 - x_i) + \prod_{i \in \mathcal{D}(\mathcal{S})} x_i \prod_{i \in \mathcal{S}} (1 - x_i) \le x_j + (1 - x_j) = 1.$$

On the other hand, if $\mathcal{S} = \emptyset$, then $\mathcal{D}(\mathcal{S}) = \Omega \setminus \{k, \ell\}$; thus, there exists $j \in \mathcal{D}(\mathcal{S})$ and

$$\prod_{i \in \mathcal{S}} x_i \prod_{i \in \mathcal{D}(\mathcal{S})} (1 - x_i) + \prod_{i \in \mathcal{D}(\mathcal{S})} x_i \prod_{i \in \mathcal{S}} (1 - x_i) \le x_j + (1 - x_j) = 1.$$

Hence,

$$\frac{\partial^2 F^M}{\partial x_\ell \partial x_k}(x) \leq \frac{D[f]}{2} \sum_{\mathcal{S}\subseteq\Omega\setminus\{k,l\}} \prod_{i\in\mathcal{S}} x_i \prod_{i\in\mathcal{D}(\mathcal{S})} (1-x_i) + \prod_{i\in\mathcal{D}(\mathcal{S})} x_i \prod_{i\in\mathcal{S}} (1-x_i)$$
$$\leq \frac{D[f]}{2}(2^{|\Omega|-2})$$
$$= 2^{|\Omega|-3} D[f].$$

Also, it is easy to observe that $\frac{\partial^2 F^M}{\partial x_k^2}(x) = 0$. Applying Proposition 5 proves the claim. □

LEMMA 2. *For any* $\gamma \in \Gamma(g_\tau), \mathcal{S} \subseteq \Omega$, *we have* $-|\Omega|D[f_\tau] + \sum_{s\in\Omega} \gamma_s(x(\mathcal{S}))_s \leq f_\tau(\mathcal{S}) - \phi(\tau)$.

Proof: Recall that $g_\tau(\emptyset) = 0$ and $D[f_\tau] = D[g_\tau]$. By Theorem 2, $-|\Omega|D[f_\tau] + \sum_{s\in\Omega} \gamma_s(x(\mathcal{S}))_s \leq G_\tau^L(x(\mathcal{S})) = g_\tau(\mathcal{S}) = f_\tau(\mathcal{S}) - \phi(\tau)$. □

PROPOSITION 6. *For any* $\gamma \in \Gamma(g_\sigma)$, *the following inequality is valid for* $\text{conv}(H_{\mathbb{B}})$*:*

$$-|\Omega|D[f_\sigma] + \sum_{s\in\Omega} \gamma_s x_s \leq z - \phi(\sigma). \tag{10}$$

Proof: This proof follows arguments similar to that of Atamtürk and Narayanan (2008). Consider $(x, z) \in H_{\mathbb{B}}$, which implies $x = x(\mathcal{S})$, for some $\mathcal{S} \subseteq \Omega$. From Lemma 2, $-|\Omega|D[f_\sigma] + \sum_{s\in\Omega} \gamma_s(x(\mathcal{S}))_s \leq f_\sigma(\mathcal{S}) - \phi(\sigma) = g_\sigma(\mathcal{S})$. Because $(x, z) \in H_{\mathbb{B}}, F_\sigma(x) = f_\sigma(\mathcal{S}) = g_\sigma(\mathcal{S}) + \phi(\sigma) \leq z$, which implies $-|\Omega|D[f_\sigma] + \sum_{s\in\Omega} \gamma_s x_s \leq z - \phi(\sigma)$. □

PROPOSITION 8. *Let* $f : 2^\Omega \to \mathbb{R}$ *be increasing with* $f(\emptyset) = 0$. *Suppose* $\gamma \in \mathbb{R}^{|\Omega|}$ *and*

$$\sum_{s\in\Omega} \gamma_s x_s \leq z + |\Omega|D[f] + \gamma_0 \tag{11}$$

defines a nontrivial facet of H_f. *Let* $\bar{f} : 2^\Omega \to \mathbb{R}$ *be defined by* $\bar{f}(\emptyset) = 0, \bar{f}(\mathcal{S}) = f(\mathcal{S}) + |\Omega|D[f] + \gamma_0$, *for all nonempty* $\mathcal{S} \subseteq \Omega$, *and suppose* $\gamma \in \Gamma(\bar{f})$. *Then,* $\gamma_0 \leq 0$.

Proof: This proof follows steps similar to that of Atamtürk and Narayanan (2020), with additional arguments to account for approximate submodularity. Suppose $\gamma_0 > 0$. Because (11) is a valid inequality for H_f, for any non-empty $\mathcal{S} \subseteq \Omega$, $\sum_{s\in\mathcal{S}} \gamma_s = \sum_{s\in\Omega} \gamma_s x(\mathcal{S})_s \leq f(\mathcal{S}) + \gamma_0 + |\Omega|D[f] = \bar{f}(\mathcal{S})$, and $\sum_{s\in\emptyset} \gamma_s = 0 = \bar{f}(\emptyset)$. Thus, $\gamma \in P_{\bar{f}}$, and by Proposition 7,

$$\sum_{s\in\Omega} \gamma_s x_s \leq z \tag{12}$$

is valid for $H_{\bar{f}}$.

We show that (12) is facet-defining for $H_{\bar{f}}$. Observe that the solutions $\{(x(\{s\}), f(\{s\}))\}_{s\in\Omega} \cup \{(x(\emptyset), 1), (x(\emptyset), 0)\}$

are $|\Omega|+2$ affinely independent solutions, so the dimension of H_f is $|\Omega|+1$. By the hypothesis, $\sum_{s\in\Omega}\gamma_s x_s \leq z+|\Omega|D[f]+\gamma_0$ is facet-defining for H_f. Thus, there exist $|\Omega|+1$ affinely independent solutions $\{(x^k,z^k)\}_{k=1}^{|\Omega|+1}$ such that $(x^k,z^k)\in H_f$ and $\sum_{s\in\Omega}\gamma_s x_s^k = z^k+|\Omega|D[f]+\gamma_0$. By Carathéodory's theorem, each of the affinely independent solutions (x^k,z^k) can be represented by a convex combination of $|\Omega|+2$ (integral) extreme points of H_f: $(x^k,z^k)=\sum_{\ell=1}^{|\Omega|+2}\lambda^{k,\ell}(x^{k,\ell},z^{k,\ell})$, where $\sum_{\ell=1}^{|\Omega|+2}\lambda^{k,\ell}=1$, $\lambda^k\in\mathbb{R}_+^{|\Omega|+2}$, and $(x^{k,\ell},z^{k,\ell})$ is an integral extreme point of H_f, for all $\ell\in\{1,\ldots,|\Omega|+2\}$. Consider $(x^{k,\ell},z^{k,\ell})$ for some $k\in\{1,\ldots,|\Omega|+1\}$ and $\ell\in\{1,\ldots,|\Omega|+2\}$. Suppose that $x^{k,\ell}=0$; because $(x^{k,\ell},z^{k,\ell})\in H_f$, $0\leq z^{k,\ell}$. Thus, $\bar{f}(x^{k,\ell})=0\leq z^{k,\ell}<z^{k,\ell}+|\Omega|D[f]+\gamma_0$. If instead $x^{k,\ell}\neq 0$, then $x^{k,\ell}=x(\mathcal{S})$ for some nonempty $\mathcal{S}\subseteq\Omega$. Notice that by $(x^{k,\ell},z^{k,\ell})\in H_f$, $f(x^{k,\ell})\leq z^{k,\ell}$, so $\bar{f}(x^{k,\ell})\leq z^{k,\ell}+|\Omega|D[f]+\gamma_0$. Hence, $(x^{k,\ell},z^{k,\ell}+|\Omega|D[f]+\gamma_0)\in H_{\bar{f}}$; moreover, $(x^k,z^k+|\Omega|D[f]+\gamma_0)=\sum_{\ell=1}^{|\Omega|+2}\lambda^{k,\ell}(x^{k,\ell},z^{k,\ell}+|\Omega|D[f]+\gamma_0)\in H_{\bar{f}}$.

Suppose that the points $(x^k,z^k+|\Omega|D[f]+\gamma_0)_{k=1}^{|\Omega|+1}$ are not affinely independent. Then there exists $\sigma\in\mathbb{R}^{|\Omega|+1}\setminus\{0\}$ such that $\sum_{k=1}^{|\Omega|+1}\sigma_k=0$ and $\sum_{k=1}^{|\Omega|+1}\sigma_k(x^k,z^k+|\Omega|D[f]+\gamma_0)=(0,0)$. Let $j\in\{1,\ldots,|\Omega|+1\}$ be such that $\sigma_j\neq 0$; without loss of generality, let $\sigma_j=1$. Thus, $\sum_{k\neq j}\sigma_k(x^k,z^k+|\Omega|D[f]+\gamma_0)=-(x^j,z^j+|\Omega|D[f]+\gamma_0)$. Because $\sigma_j=1$, $\sum_{k\neq j}\sigma_k=-1$; thus,

$$
\begin{aligned}
-(x^j,z^j+|\Omega|D[f]+\gamma_0) &= \sum_{k\neq j}\sigma_k(x^k,z^k+|\Omega|D[f]+\gamma_0)\\
&= -(0,|\Omega|D[f]+\gamma_0)+\sum_{k\neq j}\sigma_k(x^k,z^k)\\
\iff -(x^j,z^j) &= \sum_{k\neq j}\sigma_k(x^k,z^k),
\end{aligned}
$$

which contradicts the affine independence of $(x^k,z^k)_{k=1}^{|\Omega|+1}$, so $(x^k,z^k+|\Omega|D[f]+\gamma_0)_{k=1}^{|\Omega|+1}$ are affinely independent. Also, $\sum_{s\in\Omega}\gamma_s x_s^k=z^k+|\Omega|D[f]+\gamma_0$, for each $k\in\{1,\ldots,|\Omega|+1\}$, which implies (12) is facet-defining for $H_{\bar{f}}$.

By Proposition 7, γ is an extreme point of $P_{\bar{f}}$, and by the hypothesis, there exists a permutation $(\rho_1,\ldots,\rho_{|\Omega|})$ such that $\gamma_{\rho_s}=\bar{f}(\mathcal{S}_s^\rho)-\bar{f}(\mathcal{S}_{s-1}^\rho)$, for all $s\in\Omega$. Define $\hat{\gamma}$ by $\hat{\gamma}_{\rho_1}=\gamma_{\rho_1}-\gamma_0,\hat{\gamma}_{\rho_s}=\gamma_{\rho_s}$, otherwise. Thus, $\hat{\gamma}_{\rho_s}=f(\mathcal{S}_s^\rho)-f(\mathcal{S}_{s-1}^\rho)$ and $\hat{\gamma}\in\Gamma(f)$. Because $\gamma_0>0$, $\bar{f}$ is increasing, so that by Lemma 6, $\sum_{s\in\mathcal{S}}\hat{\gamma}_s\leq f(\mathcal{S})+|\mathcal{S}|D[f]$. Hence, if $\gamma'_s=\hat{\gamma}_s-D[f]$, then $\gamma'\in P_f$. By Proposition 7, we have $\sum_{s\in\Omega}\gamma'_s x_s\leq z$ is valid for H_f, which implies $(\gamma_{\rho_1}-\gamma_0)x_{\rho_1}+\sum_{s\in\Omega\setminus\{\rho_1\}}\hat{\gamma}_s x_s\leq z+|\Omega|D[f]$ is also valid for H_f. Because $\gamma_0>0$, we also have the valid inequality $\gamma_0 x_{\rho_1}\leq\gamma_0$.

Combining these last two inequalities implies

$$
\sum_{s\in\Omega}\gamma_s x_s\leq z+|\Omega|D[f]+\gamma_0,
$$

thus the facet-defining inequality (11) is dominated, a contradiction. $\square$

Proposition 17. *If $f(\{s\}) \leq b$, for all $s \in \Omega$, then X is full-dimensional.*

Proof: By the hypothesis, the zero vector and $x(\{s\})$ are feasible for each $s \in \Omega$. Hence there are $|\Omega| + 1$ affinely independent points in X, implying the dimension of X is $|\Omega|$. $\square$

Proposition 18. *(Hammer et al. 1975, Atamtürk and Narayanan 2009)*

1. *The inequality $x(\{s\}) \geq 0$ is facet-defining for $\text{conv}(X)$, for all $s \in \Omega$.*

2. *The inequality $x(\{s\}) \leq 1$ is facet-defining for $\text{conv}(X)$ if and only if $f(\{s,t\}) \leq b$ for all $t \in \Omega \backslash s$.*

PROPOSITION 9. *If $\mathcal{S} \subseteq \Omega$ is a cover for X, the extended cover inequality $\sum_{s \in E_\pi(\mathcal{S})} x_s \leq |\mathcal{S}| - 1$ is valid for X if $f(\mathcal{S}) > |\Omega| D[f] + b$. In addition, the inequality defines a facet of $\{x \in X \mid x_s = 0, \forall\ s \notin E_\pi(\mathcal{S})\}$ if $\mathcal{S}$ is also a minimal cover and for each $s \in U_\pi(\mathcal{S})$, there exist $t_s, u_s \in \mathcal{S}$ such that $t_s \neq u_s$, and $f(\mathcal{S} \cup \{s\} \backslash \{t_s, u_s\}) \leq b$.*

Proof: This proof uses steps similar to those in Atamtürk and Narayanan (2009), who establish the submodular case. We show that if $x \in [0,1]^{|\Omega|}$ with $\sum_{s \in E_\pi(\mathcal{S})} x_s > |S| - 1$, then $x_s \notin X$. Because X is the convex hull of characteristic vectors, it suffices to consider such characteristic vectors. That is $x(\tilde{S})$, where $\tilde{S} \subseteq \Omega$ and there exists $\mathcal{T} \subseteq \tilde{S}$ such that $\mathcal{T} \subseteq E_\pi(\mathcal{S})$ with $|\mathcal{T}| \geq |\mathcal{S}|$. In this case, $\sum_{s \in E_\pi(\mathcal{S})} x(\tilde{S})_s \geq \sum_{s \in \mathcal{T}} x(\tilde{S})_s \geq |\mathcal{S}|$. Let $\mathcal{K} = \mathcal{S} \backslash \mathcal{T}$, and $\mathcal{L} = U_\pi(\mathcal{S}) \cap \mathcal{T} = \{\ell_1, \dots, \ell_{|\mathcal{L}|}\}$, with indexing consistent with π.

Observe that $\mathcal{S} \backslash \mathcal{K} = \mathcal{S} \cap \mathcal{T}$ and $(\mathcal{S} \cup \mathcal{L}) \backslash \mathcal{K} = (\mathcal{S} \cup (U_\pi(\mathcal{S}) \cap \mathcal{T})) \backslash (\mathcal{S} \backslash \mathcal{T}) = (\mathcal{S} \cap \mathcal{T}) \cup (U_\pi(\mathcal{S}) \cap \mathcal{T}) = \mathcal{T}$. Hence,

$$f(\mathcal{T}) = f(\mathcal{S} \backslash \mathcal{K}) + \sum_{\ell_i \in \mathcal{L}} f((\mathcal{S} \cup \{\ell_1, \dots, \ell_i\}) \backslash \mathcal{K}) - f((\mathcal{S} \cup \{\ell_1, \dots, \ell_{i-1}\}) \backslash \mathcal{K}).$$

Given $\ell_i \in \mathcal{L}$, let $\pi_j = \ell_i$. Then $(\mathcal{S} \cup \{\ell_1, \dots, \ell_{i-1}\}) \backslash \mathcal{K} \subseteq \mathcal{S} \cup \{\pi_1, \dots, \pi_{j-1}\}$; it follows from the definition of $D[f]$ that

$$\begin{aligned} & f((\mathcal{S} \cup \{\ell_1, \dots, \ell_i\}) \backslash \mathcal{K}) - f((\mathcal{S} \cup \{\ell_1, \dots, \ell_{i-1}\}) \backslash \mathcal{K}) \\ \geq\ & f(\mathcal{S} \cup \{\pi_1, \dots, \pi_j\}) - f(\mathcal{S} \cup \{\pi_1, \dots, \pi_{j-1}\}) - D[f]. \end{aligned}$$

Therefore,

$$f(\mathcal{T}) \geq f(\mathcal{S} \backslash \mathcal{K}) - |\mathcal{L}| D[f] + \sum_{\pi_j \in \mathcal{L}} f(\mathcal{S} \cup \{\pi_1, \dots, \pi_j\}) - f(\mathcal{S} \cup \{\pi_1, \dots, \pi_{j-1}\}).$$

By the definition of $U_\pi(\mathcal{S})$, for all $s \in \mathcal{S}$,

$$f(\mathcal{S} \cup \{\pi_1, \dots, \pi_j\}) - f(\mathcal{S} \cup \{\pi_1, \dots, \pi_{j-1}\}) \geq f(\{s\}).$$

Because $\mathcal{T} = (\mathcal{S} \cup \mathcal{L}) \backslash \mathcal{K}$ and $|\mathcal{T}| \geq |\mathcal{S}|$, $|\mathcal{S} \cap \mathcal{T}| + |\mathcal{K}| = |\mathcal{S}| \leq |\mathcal{T}| = |\mathcal{S} \cap \mathcal{T}| + |\mathcal{L} \cap \mathcal{T}| = |\mathcal{S} \cap \mathcal{T}| + |\mathcal{L}|$; thus, $|\mathcal{K}| \leq |\mathcal{L}$. This implies that

$$\begin{aligned} & f(\mathcal{S} \backslash \mathcal{K}) - |\mathcal{L}| D[f] + \sum_{\pi_j \in \mathcal{L}} f(\mathcal{S} \cup \{\pi_1, \ldots, \pi_j\}) - f(\mathcal{S} \cup \{\pi_1, \ldots, \pi_{j-1}\}) \\ \geq\ & f(\mathcal{S} \backslash \mathcal{K}) - |\mathcal{L}| D[f] + \sum_{s \in \mathcal{K}} f(\{s\}). \end{aligned}$$

By the definition of $D[f]$,

$$f(\mathcal{T}) \geq f(\mathcal{S} \backslash \mathcal{K}) - (|\mathcal{L}| + |\mathcal{K}|) D[f] + \sum_{s \in \mathcal{K}} f((\mathcal{S} \backslash \mathcal{K}) \cup \{s\}) - f(\mathcal{S} \backslash \mathcal{K}).$$

By the monotonicity of f, $f(\mathcal{S}) \geq f(\mathcal{S} \backslash \mathcal{K} \cup \{s\}) \geq f(\mathcal{S} \backslash \mathcal{K})$, for all $s \in \mathcal{K}$. Also, $|\mathcal{L}| \leq |\mathcal{T}|$ and $|\mathcal{K}| \leq |\Omega| - |\mathcal{T}|$. Thus,

$$\begin{aligned} f(\mathcal{T}) &\geq f(\mathcal{S}) - |\Omega| D[f] \\ &> b, \end{aligned}$$

which follows from the hypothesis. It follows that $f(x(\tilde{\mathcal{S}})) > b$.

To prove the facet claim, observe that each of the points $x(\mathcal{S} \backslash \{s\})$, for all $s \in \mathcal{S}$ and $x(\mathcal{S} \cup \{s\} \backslash \{t_s, u_s\})$, for all $s \in U_\pi(\mathcal{S})$, are $|E_\pi(\mathcal{S})|$ affinely independent points in $\{x \in X \mid x_s = 0, \forall\ s \notin E_\pi(\mathcal{S})\}$, and the valid inequality holds with equality for these points. Thus, the valid inequality defines a facet of $\{x \in X \mid x_s = 0, \forall\ s \notin E_\pi(\mathcal{S})\}$. □

Proposition 12. *Let K be the maximum cardinality parameter and L be the number of iterations run by the greedy algorithm. Consider $\epsilon \in (0, 1)$. If f is a nonnegative, increasing, $(\mathcal{S}_L \cup \widehat{\mathcal{S}}_K, \epsilon)$-approximately submodular set function, then* $f(\mathcal{S}_L) \geq \frac{(1-\epsilon)^2 f(\widehat{\mathcal{S}}_K)}{4K\epsilon + (1-\epsilon)^2} \left(1 - \left(\frac{(K-1)(1-\epsilon)^2}{K(1+\epsilon)^2}\right)^L\right)$.

Proof: This proof follows arguments similar to those of Horel and Singer (2016). Fix L and K. Consider $\ell \in \{1, ..., L-1\}$, and let $F : 2^\Omega \to \mathbb{R}$ be a function that is submodular over $\Omega_{\mathcal{S}_L \cup \widehat{\mathcal{S}}_K}$ and $(1-\epsilon) F(\mathcal{S}) \leq f(\mathcal{S}) \leq (1+\epsilon) F(\mathcal{S})$, for all $\mathcal{S} \in \Omega_{\mathcal{S}_L \cup \widehat{\mathcal{S}}_K}$. By the local submodularity of F, the greedy algorithm, and the approximate local submodularity of f, we have

$$\begin{aligned} F(\widehat{\mathcal{S}}_K) &\leq F(\mathcal{S}_\ell) + \sum_{s \in \widehat{\mathcal{S}}_K} [F(\mathcal{S}_\ell \cup \{s\}) - F(\mathcal{S}_\ell)] \\ &\leq F(\mathcal{S}_\ell) + \sum_{s \in \widehat{\mathcal{S}}_K} \left[\frac{1}{1-\epsilon} f(\mathcal{S}_\ell \cup \{s\}) - F(\mathcal{S}_\ell)\right] \end{aligned}$$

$$
\begin{aligned}
&\leq F(\mathcal{S}_\ell) + \sum_{s \in \widehat{\mathcal{S}}_K} \left[\frac{1}{1-\epsilon} f(\mathcal{S}_{\ell+1}) - F(\mathcal{S}_\ell)\right] \\
&\leq F(\mathcal{S}_\ell) + \sum_{s \in \widehat{\mathcal{S}}_K} \left[\frac{1+\epsilon}{1-\epsilon} F(\mathcal{S}_{\ell+1}) - F(\mathcal{S}_\ell)\right] \\
&\leq F(\mathcal{S}_\ell) + K\left[\frac{1+\epsilon}{1-\epsilon} F(\mathcal{S}_{\ell+1}) - F(\mathcal{S}_\ell)\right].
\end{aligned}
$$

Rearranging the above inequality yields

$$
\begin{aligned}
&K\frac{1+\epsilon}{1-\epsilon} F(\mathcal{S}_{\ell+1}) \geq (K-1)F(\mathcal{S}_\ell) + F(\widehat{\mathcal{S}}_K) \\
\Rightarrow &\frac{K(1+\epsilon)}{(1-\epsilon)^2} f(\mathcal{S}_{\ell+1}) \geq (K-1)F(\mathcal{S}_\ell) + F(\widehat{\mathcal{S}}_K),
\end{aligned}
$$

which is equivalent to

$$
\begin{aligned}
f(\mathcal{S}_{\ell+1}) &\geq \frac{\upsilon_{K1}(1-\epsilon)^2 F(\mathcal{S}_\ell)}{(1+\epsilon)} + \frac{(1-\epsilon)^2 F(\widehat{\mathcal{S}}_K)}{K(1+\epsilon)} \\
&\geq \upsilon_{K1}\frac{(1-\epsilon)^2}{(1+\epsilon)^2} f(\mathcal{S}_\ell) + \frac{1}{K}\frac{(1-\epsilon)^2}{(1+\epsilon)^2} f(\widehat{\mathcal{S}}_K).
\end{aligned}
$$

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

$$
f(\mathcal{S}_\ell) \geq \frac{(1-\epsilon)^2 f(\widehat{\mathcal{S}}_K)}{4K\epsilon + (1-\epsilon)^2}\left(1 - \left(\frac{\upsilon_{K1}(1-\epsilon)^2}{(1+\epsilon)^2}\right)^\ell\right)
$$

and this implies

$$
f(\mathcal{S}_L) \geq \frac{(1-\epsilon)^2 f(\widehat{\mathcal{S}}_K)}{4K\epsilon + (1-\epsilon)^2}\left(1 - \left(\frac{\upsilon_{K1}(1-\epsilon)^2}{(1+\epsilon)^2}\right)^L\right). \qquad \square
$$

Lemma 9. *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 $N \in \mathbb{N}$, $a_n \leq b_n$.

$$
\begin{aligned}
a_N &\leq \alpha a_{N-1} + \beta \\
&\leq \alpha b_{N-1} + \beta \\
&= b_N.
\end{aligned}
$$

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

Lemma 10. *Let α, β, and $b_0 \in \mathbb{R}$, where $\alpha \neq 1$. Define the sequence $\{b_i\}$ by $b_{i+1} = \alpha b_i + \beta$. Then $b_i = \alpha^i \left(b_0 - \frac{\beta}{1-\alpha}\right) + \frac{\beta}{1-\alpha}$.*

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 $n = 0$:

$$\begin{aligned}\tilde{b}_0 &= \alpha^0 (b_0 - \frac{\beta}{1-\alpha}) + \frac{\beta}{1-\alpha} \\ &= b_0.\end{aligned}$$

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

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

Lemma 11. *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 $\alpha \neq 1$. Then $a_i \leq \alpha^i \left(a_0 - \frac{\beta}{1-\alpha}\right) + \frac{\beta}{1-\alpha}$.*

Proof: The proof follows by setting $b_0 = a_0$ and $b_{i+1} = \alpha b_i + \beta$ and then applying Lemmas 9 and 10. $\square$

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

THEOREM 4. *Let $K \in \{1, ..., |\Omega|\}$ be the maximum cardinality parameter and $L \in \{0, ..., |\Omega| - 1\}$ be the number of iterations run by the greedy algorithm. If f is a nonnegative, increasing set function, then*

$$f(\mathcal{S}_L) \geq \left[f(\widehat{\mathcal{S}}_K) - \min\{\Delta^{L,K}[f], f(\widehat{\mathcal{S}}_K)\}\right] \left[1 - \left(\frac{K-1}{K}\right)^L\right].$$

Moreover, the above bound is tight.

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

$$f(\widehat{\mathcal{S}}_K \cup \mathcal{S}_\ell) = f(\mathcal{S}_\ell) + \sum_{k=0}^{K-1} f(\widehat{\mathcal{S}}_K(k+1) \cup \mathcal{S}_\ell) - \sum_{k=0}^{K-1} f(\widehat{\mathcal{S}}_K(k) \cup \mathcal{S}_\ell). \tag{13}$$

For each $k \in \{0, ..., K-1\}$, using the definition of $d(\ell, k)$,

$$\begin{aligned} &f(\widehat{\mathcal{S}}_K(k+1) \cup \mathcal{S}_\ell) - f(\widehat{\mathcal{S}}_K(k) \cup \mathcal{S}_\ell) \\ \leq &f(\{\hat{s}_K^{k+1}\} \cup \mathcal{S}_\ell) - f(\mathcal{S}_\ell) + d^{\ell,k}[f]. \end{aligned} \tag{14}$$

We plug the bound obtained in (14) into (13), and use the fact that $f(\mathcal{S}_{\ell+1}) \geq f(\hat{s}_K^{k+1} \cup \mathcal{S}_\ell)$ to obtain

$$f(\widehat{\mathcal{S}}_K \cup \mathcal{S}_\ell) \leq f(\mathcal{S}_\ell) + \sum_{k=0}^{K-1} [f(\mathcal{S}_{\ell+1}) - f(\mathcal{S}_\ell) + d^{\ell,k}[f]],$$

$$\iff$$

$$f(\widehat{\mathcal{S}}_K \cup \mathcal{S}_\ell) - f(\mathcal{S}_\ell) \leq K(f(\mathcal{S}_{\ell+1}) - f(\mathcal{S}_\ell)) + \delta^{\ell,K}[f]. \tag{15}$$

Because f is increasing,

$$f(\widehat{\mathcal{S}}_K) - f(\mathcal{S}_\ell) \leq K(f(\mathcal{S}_{\ell+1}) - f(\mathcal{S}_\ell)) + \delta^{\ell,K}[f]. \tag{16}$$

Some additional arithmetic yields

$$f(\widehat{\mathcal{S}}_K) - f(\mathcal{S}_{\ell+1}) \leq \upsilon_{K1}(f(\widehat{\mathcal{S}}_K) - f(\mathcal{S}_\ell)) + \frac{\delta^{\ell,K}[f]}{K}.$$

By Lemma 11 and the nonnegativity of f, we have, for all $\ell \in \{0, ..., L\}$,

$$f(\widehat{\mathcal{S}}_K) - f(\mathcal{S}_\ell) \leq \upsilon_{K\ell}(f(\widehat{\mathcal{S}}_K) - \Delta^{L,K}[f]) + \Delta^{L,K}[f].$$

Let $\ell = L$. Then simple rearrangement of terms yields

$$\left[f(\widehat{\mathcal{S}}_K) - \Delta^{L,K}[f]\right][1 - \upsilon_{KL}] \leq f(\mathcal{S}_L), \tag{17}$$

which completes the proof of the bound's validity.

We show that there exist tight examples in which $K = L$ and $L \in \{1, ..., |\Omega| - 1\}$. Cornuéjols et al. (1977) and Fisher et al. (1978) prove that, for each such L, there exists a nonnegative, increasing, submodular function f such that

$$f(\widehat{\mathcal{S}}_L)\,[1 - \upsilon_{LL}] = f(\mathcal{S}_L). \tag{18}$$

By the proven bound, $\left[f(\widehat{\mathcal{S}}_L) - \Delta^{L,L}[f]\right][1 - \upsilon_{LL}] \leq f(\mathcal{S}_L) = f(\widehat{\mathcal{S}}_L)\,[1 - \upsilon_{LL}]$, which implies that $\Delta^{L,L}[f] \geq 0$.

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 $\ell \in \{1, ...|\Omega|\}, k \in \{0, ..., |\Omega| - 1\}$. By the definition of $\delta^{\ell,L}[f]$, we have $\delta^{\ell,L}[f] = \sum_{k=0}^{L-1} d^{\ell,k}[f] \leq 0$. This implies

$$\Delta^{L,L}[f] = \max_{\ell \in \{1,...,L\}} \delta^{\ell,L}[f] \leq 0.$$

Thus, $\Delta^{L,L}[f] = 0$, and the bound in the theorem statement is

$$f(\widehat{\mathcal{S}}_L)\,[1 - \upsilon_{LL}] \leq f(\mathcal{S}_L),$$

which we already stated is an equality in (18). □

THEOREM 5. *Let $K \in \{1, ..., |\Omega|\}$ be the maximum cardinality parameter and $L \in \{0, ..., |\Omega| - 1\}$ be the number of iterations run by the greedy algorithm. If f is a nonnegative, increasing set function, then*

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

where $\mathscr{C}_L$ is the collection of subsets made by the greedy algorithm at each iteration. Moreover, the above bound is tight.

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

$$\begin{aligned} f(\widehat{\mathcal{S}}_K \cup \mathcal{S}_\ell) = f(\mathcal{S}_\ell) &+ \sum_{k=0}^{K-1} f(\widehat{\mathcal{S}}_K(k+1) \cup \mathcal{S}_\ell) \\ &- \sum_{k=0}^{K-1} f(\widehat{\mathcal{S}}_K(k) \cup \mathcal{S}_\ell). \end{aligned} \tag{19}$$

Fix $k \in \{0, ..., K - 1\}$. By the definition of $\hat{d}^{\ell,k}[f]$,

$$f(\widehat{\mathcal{S}}_K(k+1) \cup \mathcal{S}_\ell) - f(\widehat{\mathcal{S}}_K(k) \cup \mathcal{S}_\ell)$$

$$\leq f(\{\hat{s}_K^{k+1}\} \cup \mathcal{S}_\ell) - f(\mathcal{S}_\ell) + \hat{d}^{\mathcal{S}_\ell,k}[f], \tag{20}$$

where $\mathcal{S}_\ell$ is the set chosen by the greedy algorithm after ℓ iterations. We plug the bound obtained in (20) into (19), and use the fact that $f(\mathcal{S}_{\ell+1}) \geq f(\hat{s}_K^{k+1} \cup \mathcal{S}_\ell)$ to obtain

$$f(\widehat{\mathcal{S}}_K \cup \mathcal{S}_\ell) \leq f(\mathcal{S}_\ell) + \sum_{k=0}^{K-1} [f(\mathcal{S}_{\ell+1}) - f(\mathcal{S}_\ell) + \hat{d}^{\mathcal{S}_\ell,k}[f]],$$

$$\iff$$

$$f(\widehat{\mathcal{S}}_K \cup \mathcal{S}_\ell) - f(\mathcal{S}_\ell) \leq K(f(\mathcal{S}_{\ell+1}) - f(\mathcal{S}_\ell)) + \hat{\delta}^{\mathcal{S}_\ell,K}[f]. \tag{21}$$

Because f is increasing,

$$f(\widehat{\mathcal{S}}_K) - f(\mathcal{S}_\ell) \leq K(f(\mathcal{S}_{\ell+1}) - f(\mathcal{S}_\ell)) + \hat{\delta}^{\mathcal{S}_\ell,K}[f]. \tag{22}$$

Some additional arithmetic yields

$$f(\widehat{\mathcal{S}}_K) - f(\mathcal{S}_{\ell+1}) \leq \upsilon_{K1}(f(\widehat{\mathcal{S}}_K) - f(\mathcal{S}_\ell)) + \frac{\hat{\delta}^{\mathcal{S}_\ell,K}[f]}{K}.$$

By Lemma 11 and the nonnegativity of f, we have, for all $\ell \in \{0, ..., L\}$,

$$f(\widehat{\mathcal{S}}_K) - f(\mathcal{S}_\ell) \leq \upsilon_{K\ell}(f(\widehat{\mathcal{S}}_K) - \hat{\Delta}^{\mathscr{C}_L,K}[f]) + \hat{\Delta}^{\mathscr{C}_L,K}[f].$$

Let $\ell = L$. Then simple rearrangement of terms yields

$$\left[f(\widehat{\mathcal{S}}_K) - \hat{\Delta}^{\mathcal{S}_L,K}[f]\right][1 - \upsilon_{KL}] \leq f(\mathcal{S}_L), \tag{23}$$

which completes the proof of the bound's validity. We omit a proof of the bound's tightness as it is similar to that of Theorem 4. □

Lemma 12. *Let $\mathscr{C}$ be a collection of subsets of $\mathcal{S} \subseteq \Omega$, where $\emptyset \in \mathscr{C}$, and $K \in \{0, .., |\Omega| - 1\}$ is the maximum cardinality parameter. Then $\widehat{\mathcal{I}}^{\mathscr{C},K}[f] \leq \mathcal{I}^{\mathcal{S},K}[f]$.*

Proof: This is immediate from the fact that $\mathcal{I}$ considers a superset of the $(\mathcal{A}, \mathcal{B})$ pairs of $\widehat{\mathcal{I}}$. □

Theorem 6. *Let K be the maximum cardinality parameter and L be the number of iterations run by the greedy algorithm. If f is a nonnegative, increasing set function, then $f(\mathcal{S}_L) \geq \min\{f(\widehat{\mathcal{S}}_K), (1 - (1 - 1/K)^L)[f(\widehat{\mathcal{S}}_K) - \min\{\widehat{\mathcal{I}}(\mathscr{C}_L, K), f(\widehat{\mathcal{S}}_K)\}]\}$.*

Proof: Suppose $|\widehat{\mathcal{S}}_K \backslash \mathcal{S}_{L-1}| \geq 2$, which implies that $|\widehat{\mathcal{S}}_K \backslash \mathcal{S}_\ell| \geq 2$, for all $\ell \in \{0, ..., L-1\}$. Fix $\ell \in$

$\{0, ..., L-1\}$. Observe that

$$K(f(\mathcal{S}_{\ell+1}) - f(\mathcal{S}_\ell)) = K \sum_{s \in \mathcal{S}_{\ell+1} \setminus \mathcal{S}_\ell} (f(\mathcal{S}_\ell \cup \{s\}) - f(\mathcal{S}_\ell))$$
$$\geq \sum_{s \in \widehat{\mathcal{S}}_K \setminus \mathcal{S}_\ell} (f(\mathcal{S}_\ell \cup \{s\}) - f(\mathcal{S}_\ell)). \tag{24}$$

By the definition of the local submodularity index (Definition 1), the right-hand side of (24) can be rewritten as

$$-\phi^{\mathcal{S}_\ell, \widehat{\mathcal{S}}_K \setminus \mathcal{S}_\ell}[f] + f(\widehat{\mathcal{S}}_K \cup \mathcal{S}_\ell) - f(\mathcal{S}_\ell). \tag{25}$$

Hence,

$$f(\widehat{\mathcal{S}}_K \cup \mathcal{S}_\ell) - f(\mathcal{S}_\ell) \leq K(f(\mathcal{S}_{\ell+1}) - f(\mathcal{S}_\ell)) + \widehat{\mathcal{I}}^{\mathscr{C}_L, |\widehat{\mathcal{S}}_K \setminus \mathcal{S}_\ell|}[f].$$

The function $f(\cdot)$ is increasing; therefore,

$$f(\widehat{\mathcal{S}}_K) - f(\mathcal{S}_\ell) \leq K(f(\mathcal{S}_{\ell+1}) - f(\mathcal{S}_\ell)) + \widehat{\mathcal{I}}^{\mathscr{C}_L, |\widehat{\mathcal{S}}_K \setminus \mathcal{S}_\ell|}[f],$$

which is equivalent to

$$f(\widehat{\mathcal{S}}_K) - f(\mathcal{S}_{\ell+1}) \leq \upsilon_{K1}(f(\widehat{\mathcal{S}}_K) - f(\mathcal{S}_\ell)) + \frac{\widehat{\mathcal{I}}^{\mathscr{C}_L, |\widehat{\mathcal{S}}_K \setminus \mathcal{S}_\ell|}[f]}{K}.$$

From this, it follows that

$$f(\widehat{\mathcal{S}}_K) - f(\mathcal{S}_{\ell+1}) \leq \upsilon_{K1}(f(\widehat{\mathcal{S}}_K) - f(\mathcal{S}_\ell)) + \frac{\widehat{\mathcal{I}}^{\mathscr{C}_L, K}[f]}{K}.$$

The last inequality uses the fact that $\widehat{\mathcal{I}}^{\mathscr{C}_L, J}[f] \leq \widehat{\mathcal{I}}^{\mathscr{C}_L, K}[f]$ for all $J \leq K$ with $|\widehat{\mathcal{S}}_K \setminus \mathcal{S}_J| \geq 2$. By Lemma 11 and the nonnegativity of f,

$$f(\widehat{\mathcal{S}}_K) - f(\mathcal{S}_\ell) \leq \upsilon_{K\ell}(f(\widehat{\mathcal{S}}_K) - \widehat{\mathcal{I}}^{\mathscr{C}_L, K}[f]) + \widehat{\mathcal{I}}^{\mathscr{C}_L, K}[f].$$

When $\ell = L$,

$$f(\widehat{\mathcal{S}}_K) - f(\mathcal{S}_L) \leq \upsilon_{KL} \left[f(\widehat{\mathcal{S}}_K) - \widehat{\mathcal{I}}^{\mathscr{C}_L, K}[f] \right] + \widehat{\mathcal{I}}^{\mathscr{C}_L, K}[f] \tag{26}$$
$$\iff f(\mathcal{S}_L) \geq (1 - \upsilon_{KL}) \left[f(\widehat{\mathcal{S}}_K) - \widehat{\mathcal{I}}^{\mathscr{C}_L, K}[f] \right]. \tag{27}$$

If $\min\{\widehat{\mathcal{I}}^{\mathscr{C}_L,K}[f], f(\widehat{\mathcal{S}}_K)\} = \widehat{\mathcal{I}}^{\mathscr{C}_L,K}[f]$, then $f(\mathcal{S}_L) \geq (1-\upsilon_{KL})\left[f(\widehat{\mathcal{S}}_K) - \widehat{\mathcal{I}}^{\mathscr{C}_L,K}[f]\right]$, by (27). If instead, $\min\{\widehat{\mathcal{I}}^{\mathscr{C}_L,K}[f], f(\widehat{\mathcal{S}}_K)\} = f(\widehat{\mathcal{S}}_K)$, then $f(\mathcal{S}_L) \geq 0 = (1-\upsilon_{KL})\left[f(\widehat{\mathcal{S}}_K) - f(\widehat{\mathcal{S}}_K)\right]$.

Now, suppose that $|\widehat{\mathcal{S}}_K \backslash \mathcal{S}_{L-1}| \leq 1$. If $\widehat{\mathcal{S}}_K \subseteq \mathcal{S}_{L-1}$, then by the monotonicity of $f(\cdot)$, $f(\mathcal{S}_L) \geq f(\widehat{\mathcal{S}}_K)$. Otherwise, let $\{s\} = \widehat{\mathcal{S}}_K \backslash \mathcal{S}_L$ and $\{t\} = \mathcal{S}_L \backslash \mathcal{S}_{L-1}$. By the greedy algorithm and monotonicity, $f(\mathcal{S}_L) = f(\mathcal{S}_{L-1} \cup \{t\}) \geq f(\mathcal{S}_{L-1} \cup \{s\}) \geq f(\widehat{\mathcal{S}}_K)$. Therefore,

$$f(\mathcal{S}_L) \geq \min\left\{f(\widehat{\mathcal{S}}_K), (1-\upsilon_{KL})\left[f(\widehat{\mathcal{S}}_K) - \min\{\widehat{\mathcal{I}}^{\mathscr{C}_L,K}[f], f(\widehat{\mathcal{S}}_K)\}\right]\right\}. \qquad \square$$

THEOREM 7. *Assume the conditions of Proposition* 10 *are satisfied, and* $f(\mathcal{S}_K) > 0$. *Then the bound from Theorem* 6 *is tighter than the bound from Proposition* 10.

Proof: Observe:

$$\begin{aligned}
&(1-\upsilon_{KK})\left[f(\widehat{\mathcal{S}}_K) - \min\{\widehat{\mathcal{I}}^{\mathcal{S}_K,K}[f], f(\widehat{\mathcal{S}}_K)\}\right] \\
&\geq (1-\frac{1}{e})\left[f(\widehat{\mathcal{S}}_K) - \min\{\mathcal{I}^{\mathcal{S}_K,K}[f], f(\widehat{\mathcal{S}}_K)\}\right] \\
&\geq \left(1 - \frac{1}{e} - \frac{\min\{\mathcal{I}^{\mathcal{S}_K,K}[f], f(\widehat{\mathcal{S}}_K)\}}{f(\widehat{\mathcal{S}}_K)}\right) f(\widehat{\mathcal{S}}_K) \\
&\geq \left(1 - \frac{1}{e} - \frac{\mathcal{I}^{\mathcal{S}_K,K}[f]}{f(\widehat{\mathcal{S}}_K)}\right) f(\widehat{\mathcal{S}}_K) \\
&\geq \left(1 - \frac{1}{e} - \frac{\mathcal{I}^{\widehat{\mathcal{S}}_K,K}[f]}{f(\mathcal{S}_K)}\right) f(\widehat{\mathcal{S}}_K). \qquad \square
\end{aligned}$$

Lemma 13. *Let* K *be the maximum cardinality parameter, and* $\mathcal{A} \subseteq \Omega$, *with* $|\mathcal{A}| = \ell$. *Then* $\phi^{\mathcal{A},\mathcal{B}}[f] \leq \hat{\delta}^{\mathcal{A},K}[f]$, *for any* $\mathcal{A} \subseteq \Omega$, *with* $|\mathcal{B}| = K$.

Proof: Let $\mathcal{B}(k) = \{b_1, ..., b_k\}$, for each $k \in \{0, ..., K\}$. Observe:

$$\begin{aligned}
\phi^{\mathcal{A},\mathcal{B}}[f] =& f(\mathcal{A} \cup \mathcal{B}) - f(\mathcal{A}) \\
&- \sum_{b \in \mathcal{B}} [f(\mathcal{A} \cup \{b\}) - f(\mathcal{A})] \\
=& \sum_{k=0}^{K-1} [f(\mathcal{A} \cup \mathcal{B}(k+1)) - f(\mathcal{A} \cup \mathcal{B}(k))] \\
&- \sum_{k=0}^{K-1} [f(\mathcal{A} \cup \{b_{k+1}\}) + f(\mathcal{A})] \\
\leq& \sum_{k=0}^{K-1} \hat{d}^{\mathcal{A},k}[f] \\
=& \hat{\delta}^{\mathcal{A},K}[f]. \qquad \square
\end{aligned}$$

Proposition 14. *We have* $D[f] \le p||w||_1^{p-1}||w||_\infty$.

Proof: Let $\mathcal{W} = \{s \in \Omega \mid w_s > 0\}$. Let $\ell \in \{0, \ldots, |\Omega| - 1\}, k \in \{0, \ldots, |\Omega|\}$, $\mathcal{A}, \mathcal{B} \subseteq \Omega, s \in \Omega, |\mathcal{A}| = \ell, |\mathcal{B}| = k$. First, suppose that $s \notin \mathcal{W}$. Then $(f(\mathcal{A} \cup \mathcal{B} \cup \{s\}) - f(\mathcal{A} \cup \mathcal{B})) - (f(\mathcal{A} \cup \{s\}) - f(\mathcal{A})) = u_s - u_s = 0 \le p||w||_1^{1-p}||w||_\infty$.

Next, suppose $s \in \mathcal{W}$. Observe that H is Lipschitz continuous on $[0, ||w||_1]$, the codomain of G: for any $z_1, z_2 \in [0, ||w||_1]$, we have $|H(z_2) - H(z_1)| \le ||H'||_\infty |z_2 - z_1| \le p||w||_1^{p-1}|z_2 - z_1|$. Hence, $f(\mathcal{A} \cup \mathcal{B} \cup \{s\}) - f(\mathcal{A} \cup \mathcal{B}) \le u_s + p||w||_1^{p-1} w_s \le u_s + p||w||_1^{p-1}||w||_\infty$. Also, $f(\mathcal{A} \cup \{s\}) - f(\mathcal{A}) \ge u_s$. Therefore, $(f(\mathcal{A} \cup \mathcal{B} \cup \{s\}) - f(\mathcal{A} \cup \mathcal{B})) - (f(\mathcal{A} \cup \{s\}) - f(\mathcal{A})) \le p||w||_1^{p-1}||w||_\infty$, which implies $D[f] \le p||w||_1^{p-1}||w||_\infty$.
□

Proposition 15. *Given an instance of CUFLP, we have* $d^{\ell,k}[h] \le |\mathrm{supp}(u)| \max\{u_{pq} \mid (p,q) \in \Omega^2\}$ *for all* $\ell \in \{0, ..., m-1\}, k \in \{0, ..., m\}$.

Proof: Because u is nonnegative, $f(\mathcal{S}) \le h(\mathcal{S})$ for all $\mathcal{S} \subseteq \Omega$. Further, $f(\mathcal{S}) \ge h(\mathcal{S}) - |\mathrm{supp}(u)| \max\{u_{pq} \mid (p,q) \in \Omega^2\}$. Let $\mathcal{A}, \mathcal{B} \subseteq \Omega, s \in \Omega \backslash \mathcal{A}$, where $|\mathcal{A}| = \ell \in \{0, ..., m-1\}$ and $|B| = k \in \{0, ..., m\}$.

$$
\begin{aligned}
& h(\mathcal{A} \cup \mathcal{B} \cup \{s\}) - h(\mathcal{A} \cup \mathcal{B}) - h(\mathcal{A} \cup \{s\}) + h(\mathcal{A}) \\
\le\ & h(\mathcal{A} \cup \mathcal{B} \cup \{s\}) - f(\mathcal{A} \cup \mathcal{B}) - f(\mathcal{A} \cup \{s\}) + h(\mathcal{A}) \\
\le\ & f(\mathcal{A} \cup \mathcal{B} \cup \{s\}) - f(\mathcal{A} \cup \mathcal{B}) - f(\mathcal{A} \cup \{s\}) + f(\mathcal{A}) \\
& + |\mathrm{supp}(u)| \max\{u_{pq} \mid (p,q) \in \Omega^2\} \\
\le\ & |\mathrm{supp}(u)| \max\{u_{pq} \mid (p,q) \in \Omega^2\}.
\end{aligned}
$$

It follows immediately that $d^{\ell,k}[f] \le |\mathrm{supp}(u)| \max\{u_{pq} \mid (p,q) \in \Omega^2\}$. □

Proposition 16. *For the above instances of cooperative facility location problems, with* f *as the submodular function such that* $(1 - \epsilon_H) f(\mathcal{S}) \le h(\mathcal{S}) \le (1 + \epsilon_H) f(\mathcal{S})$*, the smallest valid* ϵ_H *in Proposition* 12 *is* $u_{6,7}/f(6,7)$.

Proof: For ϵ to be valid for Proposition 12, $(1-\epsilon) f(\mathcal{S}) \le h(\mathcal{S}) \le (1+\epsilon) f(\mathcal{S})$, for all $\mathcal{S} \subseteq \Omega$. Because $u_{pq} \ge 0$, for all $(p,q) \in \Omega^2$, $f(\mathcal{S}) \le h(\mathcal{S})$, for all $\mathcal{S} \subseteq \Omega$. Because $\epsilon_H > 0$ and f is nonnegative, it satisfies

$$
h(\mathcal{S}) \ge f(\mathcal{S}) - \epsilon_H f(\mathcal{S}), \forall\ \mathcal{S} \subseteq \Omega.
$$

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

$$
h(\mathcal{S}) = f(\mathcal{S}) + u_{6,7}
$$

$$C^4 = \begin{bmatrix} 48 & 0 & 0 & 64 & 0 & 0 & 0 \\ 48 & 0 & 0 & 0 & 64 & 0 & 0 \\ 48 & 0 & 0 & 0 & 0 & 64 & 0 \\ 48 & 0 & 0 & 0 & 0 & 0 & 64 \\ 0 & 36 & 0 & 64 & 0 & 0 & 0 \\ 0 & 36 & 0 & 0 & 64 & 0 & 0 \\ 0 & 36 & 0 & 0 & 0 & 64 & 0 \\ 0 & 36 & 0 & 0 & 0 & 0 & 64 \\ 0 & 0 & 27 & 64 & 0 & 0 & 0 \\ 0 & 0 & 27 & 0 & 64 & 0 & 0 \\ 0 & 0 & 27 & 0 & 0 & 64 & 0 \\ 0 & 0 & 27 & 0 & 0 & 0 & 64 \end{bmatrix}$$

Figure 3: An instance of an uncapacitated facility location problem from Cornuéjols et al. (1977). The rows of C^4 correspond to clients and the columns correspond to facilities. The entry C^4_{ij} is the profit produced when facility j fulfills the demand of client i.

$$\begin{aligned} &\leq f(\mathcal{S}) + \frac{u_{6,7}}{f(\{6,7\})} f(\mathcal{S}) \\ &= f(\mathcal{S}) + \epsilon_H f(\mathcal{S}). \end{aligned}$$

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

B Examples

Example 1. *We modify the uncapacitated facility location problem described by Figure 3 to generate instances for the cooperative uncapacitated facility location problem (Section 6.2). In particular we add cooperative bonuses.*

Example 2. *Consider* $\Omega = \{1,2,3,4,5,6\}$ *and the following nonnegative, monotone set functions* $f, g : 2^\Omega \to \mathbb{R}$ *defined by*

$$f(\mathcal{S}) = \begin{cases} |\mathcal{S}|^2, & \text{if } \{2,3\} \subseteq \mathcal{S} \text{ or } \mathcal{S} \cap \{4,5,6\} \neq \emptyset \\ 0, & \text{otherwise.} \end{cases}$$

$$g(\mathcal{S}) = \begin{cases} |\mathcal{S}|^2, & \text{if } \mathcal{S} = \{4,5,6\} \text{ or } \mathcal{S} \cap \{1,2,3\} \neq \emptyset \\ 0, & \text{otherwise.} \end{cases}$$

Consider $\mathcal{T} = \{1,4\}$, $K = 2$. *By choosing* $\mathcal{A} = \{1\}, \mathcal{B} = \{2,3\}$, *we have* $\hat{\gamma}^{\mathcal{T},K}[f]\big(f(\{1,2,3\}) - f(\{1\})\big) = 9\hat{\gamma}^{\mathcal{T},K}[f] \leq f(\{1,2\}) - f(\{1\}) + f(\{1,3\}) - f(\{1\}) = 0$, *implying* $\hat{\gamma}^{\mathcal{T},K}[f] = 0$. *By choosing* $\mathcal{A} = \{4\}, \mathcal{B} = \{5,6\}$, *we have* $\hat{\gamma}^{\mathcal{T},K}[g]\big(g(\{4,5,6\}) - g(\{4\})\big) = 9\hat{\gamma}^{\mathcal{T},K}[g] \leq g(\{4,5\}) - g(\{4\}) + g(\{4,6\}) - g(\{4\}) = 0$, *implying* $\hat{\gamma}^{\mathcal{T},K}[g] = 0$.

To show that $\hat{\gamma}^{\mathcal{T},K}[f+g] > 0$, *we first note that if* $|\mathcal{B}| \leq 1$, *then* $(f+g)(\mathcal{A} \cup \mathcal{B}) - (f+g)(\mathcal{A}) = \sum_{s \in \mathcal{B}} (f+g)(\mathcal{A} \cup \{s\}) - (f+g)(\mathcal{A})$, *thus, these constraints cannot force* $\hat{\gamma}^{\mathcal{T},K}[f+g]$ *equal to 0.*

Nonnegativity and monotonicity of f and g imply that it is sufficient to show that $\sum_{s \in \mathcal{B}} (f+g)(\mathcal{A} \cup \{s\}) - (f+g)(\mathcal{A}) > 0$, for all $\mathcal{A} \subseteq \mathcal{T}, \mathcal{B} \subseteq \Omega, \mathcal{A} \cap \mathcal{B} = \emptyset, |\mathcal{B}| = 2$.

($\mathcal{A} = \emptyset$): We have $f(\mathcal{A}) = g(\mathcal{A}) = 0$. Note that either $\mathcal{B} \cap \{1,2,3\} \neq \emptyset$ or $\mathcal{B} \cap \{4,5,6\} \neq \emptyset$. Hence, for some $s \in \mathcal{B}$, either $f(\mathcal{A} \cup \{s\}) = 1$ or $g(\mathcal{A} \cup \{s\}) = 1$, which implies $\sum_{s \in \mathcal{B}} (f+g)(\mathcal{A} \cup \{s\}) - (f+g)(\mathcal{A}) \geq 1$.

($\mathcal{A} = \{1\}$): Observe that $f(\mathcal{A}) = 0, g(\mathcal{A}) = 1$. In addition, either $\mathcal{B} = \{2,3\} \neq \emptyset$ or $\mathcal{B} \cap \{4,5,6\} \neq \emptyset$. Thus, for some $s \in \mathcal{B}$, either $f(\mathcal{A} \cup \{s\}) = 4$ or $g(\mathcal{A} \cup \{s\}) = 4$, which implies $\sum_{s \in \mathcal{B}} (f+g)(\mathcal{A} \cup \{s\}) - (f+g)(\mathcal{A}) \geq 2$.

($\mathcal{A} = \{4\}$): Observe that $g(\mathcal{A}) = 0, f(\mathcal{A}) = 1$. In addition, either $\mathcal{B} \cap \{1,2,3\} \neq \emptyset$ or $\mathcal{B} \cap \{4,5,6\} \neq \emptyset$. Thus, for some $s \in \mathcal{B}$, either $f(\mathcal{A} \cup \{s\}) = 4$ or $g(\mathcal{A} \cup \{s\}) = 4$, which implies $\sum_{s \in \mathcal{B}} (f+g)(\mathcal{A} \cup \{s\}) - (f+g)(\mathcal{A}) \geq 2$.

($\mathcal{A} = \{1,4\}$): We have $f(\mathcal{A}) = 4 = g(\mathcal{A})$. For all $\mathcal{B}$ satisfying the requirements, and for any $s \in \mathcal{B}$, $f(\mathcal{A} \cup \{s\}) = 9$ because $4 \in \mathcal{A} \cup \{s\}$ and $g(\mathcal{A} \cup \{s\}) = 9$ because $1 \in \mathcal{A} \cup \{s\}$. Thus, $\sum_{s \in \mathcal{B}} (f+g)(\mathcal{A} \cup \{s\}) - (f+g)(\mathcal{A}) = 10$.

Thus, we can determine that $\hat{\gamma}^{\mathcal{T},K}[f+g] > 0 = \hat{\gamma}^{\mathcal{T},K}[f] + \hat{\gamma}^{\mathcal{T},K}[g]$, thus the submodularity ratio is not subadditive.

Example 3. *To illustrate Remark 3, consider an instance of the cooperative uncapacitated facility location problem in which $m = 3, n = 1$, and $v_{i1} = 0$, for $i = 1,2,3$, $b_1 = 1$, $u_{2,3} = 1$, and $u_{pq} = 0$ otherwise. The fixed costs are zero so h is increasing. Consider $\mathcal{A} = \{1\}, \mathcal{B} = \{2\}$, and $s = \{3\}$. Then,*

$$
\begin{aligned}
& h(\mathcal{A} \cup \mathcal{B} \cup \{s\}) = 1, \quad h(\mathcal{A} \cup \mathcal{B}) = 0, \\
& h(\mathcal{A} \cup \{s\}) = 0, \quad \textit{and} \quad h(\mathcal{A}) = 0 \\
& \Rightarrow 1 = h(\mathcal{A} \cup \mathcal{B} \cup \{s\}) - h(\mathcal{A} \cup \mathcal{B}) - h(\mathcal{A} \cup \{s\}) + h(\mathcal{A}).
\end{aligned}
$$

By Lemma 1, h is not submodular.