Abstract

The seminal work by Edmonds [9] and Lovász [39] shows the strong connection between submodular functions and convex functions. Submodular functions have tight modular lower bounds, and a subdifferential structure [16] in a manner akin to convex functions. They also admit polynomial time algorithms for minimization and satisfy the Fenchel duality theorem [18] and the Discrete Separation Theorem [15], both of which are fundamental characteristics of convex functions. Submodular functions also show signs similar to concavity. Submodular function maximization, though NP hard, admits constant factor approximation guarantees. Concave functions composed with modular functions are submodular, and they also show the diminishing returns property. In this manuscript, we try to provide a more complete picture on the relationship between submodularity with convexity and concavity, by extending many of the results connecting submodularity with convexity [39, 15, 18, 9, 16] to the concave aspects of submodular functions.

We first show the existence of the superdifferentials and efficiently computable tight modular upper bounds of a submodular function. While we show that it is hard to characterize this polyhedron, we obtain inner and outer bounds on the superdifferential along with certain specific and useful supergradients. We then investigate forms of concave extensions of submodular functions and show interesting relationships to submodular maximization. We next show connections between optimality conditions over the superdifferentials and submodular maximization, and show how forms of approximate optimality conditions translate into approximation factors for maximization. We end this paper by studying versions of the discrete separation theorem and the Fenchel duality theorem when seen from the concave point of view. In every case, we relate our results to the existing results from the convex point of view, thereby improving the analysis of the relationship between submodularity, convexity, and concavity.

1 Introduction

Long known to be an important property for problems in combinatorial optimization, economics, operations research, and game theory, submodularity is gaining popularity in a number of new areas including machine learning. Along with its natural connection to many application domains, it also admits a number of interesting theoretical characterizations. A function $f : 2^V \rightarrow \mathbb{R}$ over a ground set $V = \{1, 2, \cdots, n\}$ is submodular if for all subsets $S, T \subseteq V$, it holds that,

$$f(S) + f(T) \geq f(S \cup T) + f(S \cap T).$$  \hspace{1cm} (1)
Equivalently, a submodular set function satisfies *diminishing marginal returns*: Let $f(j|S) = f(S \cup \{j\}) - f(S)$ denote the marginal cost of element $j \in V$ with respect to $S \subseteq V$. The diminishing returns property states that,

$$f(j|S) \geq f(j|T), \forall S \subseteq T \text{ and } j \notin T. \tag{2}$$

Through the rest of the paper below, we shall also assume without loss of generality that $f(\emptyset) = 0$.

**Submodularity and convexity:** Submodular functions have been strongly associated with convex functions, to the extent that submodularity is sometimes regarded as a discrete analogue of convexity [19]. This relationship is evident by the fact that submodular function minimization is easy, in that there exist strongly polynomial time algorithms which achieve it. This is akin to convex minimization which is also easy. A number of recent results, however, make this relationship much more formal. For example, similar to convex functions, submodular functions have tight modular lower bounds and admit a sub-differential characterization [16]. Moreover, it is possible [18] to provide optimality conditions, in a manner analogous to the Karush-Kuhn-Tucker (KKT) conditions from convex programming, for submodular function minimization. Furthermore, the Fenchel duality theorem and the discrete separation theorem, both of which are known to hold for convex functions have been shown to go through for submodular functions [18, 15]. In addition, submodular functions also admit a natural convex extension, known as the Lovász extension, which is easy to evaluate [39] and optimize. The Lovász extension, moreover, also has no integrality gap and minimizing a submodular function is equivalent to minimizing its Lovász extension. All these results show that submodularity is indeed extremely closely related to convexity, and seem to verify the claim that submodularity is the discrete analog of convexity.

**Submodular functions and Concavity:** Submodular functions also have some properties, which are unlike convexity, but perhaps more akin to concavity. Submodular function maximization is known to be NP hard. However, there exist a number of constant factor approximation algorithms based on simple greedy or local search heuristics [11, 36, 43] and some recent continuous approximation methods [6, 12]. This is unlike convexity where maximization can be hopelessly difficult [45]. Furthermore, submodular functions have a diminishing returns property which is akin to concavity, and concave over modular functions are known to be submodular. In addition submodular function has been shown to have tight modular upper bounds [30, 26, 25, 32, 34], and as we show, form superdifferentials and supergradients like concave functions. The multi-linear extension of a submodular function, which has become useful recently [6] in the context of submodular maximization, is known to be concave when restricted to a particular direction. All these seem to indicate that submodular functions are related to both convexity and concavity, and have some strange properties enabling them to get the best of both classes of functions. We formalize all these relationships in this paper.

---

1 We also use this notation for sets $A, B$ as in $f(A|B) = f(A \cup B) - f(B)$. 

---
1.1 Motivation and Past Work

Since almost four decades, researchers have been investigating several theoretical and algorithmic aspects relating to submodular functions. Most of the bulk of this work [19, 9, 39, 16, 15, 17], has been in relating submodular functions to convexity from a polyhedral perspective, thereby enabling efficient algorithms for submodular minimization. From a polyhedral perspective, Fujishige, Edmonds and others [19, 9, 16], provided a characterization of the submodular polyhedron, the base polytope, and sub-differentials of a submodular functions. Thanks to these results, Lovász [39] provided an efficient characterization of the convex extension of a submodular function, which is called the Lovász extension. Finally, the connection between submodularity and convexity was made more precise when it was shown that [15, 17], that the Discrete Separation Theorem, Fenchel duality Theorem, and the Minkowski Sum theorems hold for submodular functions, when seen in analogy to convexity. From an algorithmic perspective, these results have provided several algorithms for submodular function minimization. In particular, [19, 2] use the submodular polyhedron and the convex extension to provide an exact algorithm for submodular minimization. Similarly, [24, 23] and others have used many of these ideas to provide exact algorithms for submodular minimization.

While submodular functions have been seen to be related to concavity (as discussed above), the polyhedral aspects of submodular functions from a maximization perspective are much lesser understood. Most work here has been in relation to approximation algorithms for submodular maximization. The first set of results for submodular maximization were shown in [43, 42], where they provide a $1 - 1/e$ approximation algorithm (in the form of a simple greedy heuristic) for maximizing a monotone submodular function under a cardinality constraint. Further variants of the greedy algorithm were also extended to matroid and knapsack constraints [14, 47, 35, 38]. The factor $1 - 1/e$ was also shown to be optimal under the value oracle model [10, 42]. The first systematic study for non-monotone submodular function maximization was performed by Fiege et al [11], where they obtain a $1/3$ and a randomized $2/5$ approximation for unconstrained submodular maximization. They also show an absolute hardness of $1/2$ for this problem. They raise an open question, however, whether there exists a tight $1/2$ approximation algorithm for this problem. This question was resolved in [4], where they show that a simple randomized linear time algorithm achieves an approximation factor of $1/2$. Many of these results can be extended to matroid and knapsack constraints in [50, 37].

Polyhedral aspects of submodular maximization, and the concave extension of a submodular function has been studied only in a limited context in previous work [11, 5, 49, 8, 43, 26, 25, 34, 9]. For example, a recent chain of work by Jan Vondrák and others [5, 49, 8] investigated concave extensions of a submodular function, which were shown to be NP hard to evaluate [49]. Similarly the semidifferentials have gained a lot of attention from the machine learning community. In particular, the subgradients and supergradients of a submodular function have inspired a unifying Majorization-Minimization framework for submodular optimization [41, 26, 25, 34, 27, 29]. These semidifferentials have also been used in the context of approximate inference in a class of probability distributions defined via submodular functions [28, 7], and have also been used to define a class of Bregman divergences using submodular functions [26].

In this paper, we attempt to provide a first unifying characterization of the concave aspects of submodular functions from a polyhedral perspective, thereby extending many of the observations made in [39]. In this effort, we discover a number of interesting connections between these different aspects of submodular functions connecting concavity, and contrast them to known results of submodularity and convexity.
1.2 Our Contributions

The main contributions of this work is in providing the first systematic theoretical study related to polyhedral aspects of submodular function maximization and connections to concavity. The following provides a summary of the main components of this paper.

- We show that submodular functions have tight modular (additive) upper bounds, thereby proving the existence of the superdifferential of a submodular function. We show that characterizing this superdifferential is NP hard in general. However, we provide a series of (successively tighter) outer and also inner polyhedral bounds, all obtainable in polynomial time, and also show that we can obtain some specific practically useful supergradients in polynomial time. Along the way, we relate this to $M^\natural$-concave submodular functions \cite{40} defined on $2^V$.

- We also extend the notion of the (lower) submodular polyhedron (which consists of modular lower bounds of a submodular function), and define the submodular upper polyhedron (which consists of the modular upper bounds of the submodular function).

- We define the concave extension of a submodular function, in a manner similar to the convex extension, as a linear program over the submodular upper polyhedron. We show that this is identical to the concave extensions considered in the past \cite{5,49}. We also provide a family of concave extensions, some of which can be efficiently computed in polynomial time, and finally relate these extensions to submodular function maximization.

- We then show how we can define forms of optimality conditions for submodular maximization through the submodular superdifferential. We also show how optimality conditions related to approximations to the superdifferential lead to a number of familiar approximation guarantees for these problems.

- Finally we study the Fenchel duality and Discrete Separation theorems for submodular functions seen in connection to concavity. While in general this does not hold, we show that these hold under certain mild conditions. We also show how the Minkowski-Sum theorem also holds under certain restricted conditions.

- Throughout this paper, we point interesting connections, how our results generalize many of the results of $M^\natural$-concave submodular functions \cite{40}, where many of these characterizations are exact.

1.3 Road-Map of this paper

In Sections 2, 3, 4, and 5 we review the connections between submodularity and convexity. Most of the results in these sections are from \cite{39,19}, and in some cases we provide some generalizations. In Section 2, we review polyhedral aspects of submodularity and convexity, and investigate the submodular polyhedron, submodular subdifferentials etc. In Section 3, we study the convex extensions of a submodular function, while in Section 4 we review the optimality conditions of submodular function minimization from a polyhedral perspective. Finally, in Section 5 we review the Discrete Separation Theorem, Fenchel Duality Theorem, and Minkowski Sum Theorem, all from the perspective of the convex analogy of submodular functions. In Section 6 we investigate the polyhedral aspects of submodularity and concavity, by defining the submodular upper polyhedron, and the submodular superdifferentials. In Section 7 we provide a characterization of the concave extension of a submodular function. In Section 8 we study the optimality conditions of submodular function maximization from a polyhedral perspective. Finally, in Section 9 we provide versions of the Discrete Separation
Theorem, Fenchel Duality Theorem and the Minkowski Sum Theorem, from the perspective of concavity of a submodular function.

2 Polyhedral aspects of Submodularity and Convexity

Most of the results in this section are covered in [39, 19] and the references contained therein, so for more details please refer to these texts. We use this section to review existing work on the polyhedral connections between submodularity and convexity, and to help contrast these with the corresponding results on the polyhedral connections between submodularity and concavity starting in Section 6.

2.1 Submodular (Lower) Polyhedron

For a submodular function $f$, the (lower) submodular polyhedron $P_f$ and the base polytope $B_f$ of a submodular function [19] are defined, respectively, as:

$$P_f = \{ x : x(S) \leq f(S), \forall S \subseteq V \}, \quad B_f = P_f \cap \{ x : x(V) = f(V) \}. \quad (3)$$

where $x(S) = \sum_{i \in S} x_i$. The submodular polyhedron has a number of interesting properties. An important property of the polyhedron is that the extreme points and facets can easily be characterized even though the polyhedron itself is described by an exponential number of inequalities. In fact, surprisingly, every extreme point of the submodular polyhedron is an extreme point of the base polytope. These extreme points admit an interesting characterization in that they can be computed via a simple greedy algorithm — let $\sigma$ be a permutation of $V = \{1, 2, \cdots, n\}$. Each such permutation defines a chain with elements $S_0^\sigma = \emptyset$, $S_i^\sigma = \{ \sigma(1), \sigma(2), \ldots, \sigma(i) \}$ such that $S_0^\sigma \subseteq S_1^\sigma \subseteq \cdots \subseteq S_n^\sigma$. This chain defines an extreme point $h^\sigma$ of $P_f$ with entries

$$h^\sigma(\sigma(i)) = f(S_i^\sigma) - f(S_{i-1}^\sigma). \quad (4)$$

Each permutation of $V$ characterizes an extreme point of $P_f$ and all possible extreme points of $P_f$ can be characterized in this manner [19]. Furthermore, the problem $\max_{y \in P_f} y^\top x$, which is a linear program over a submodular polyhedron, can be very efficiently computed through the greedy algorithm [9]. The following lemma gives the greedy algorithm for finding this.

2Since the submodular polyhedron consists of modular lower bounds of a submodular function, we shall also call it the lower submodular polyhedron to contrast with the submodular upper polyhedron we introduce in Section 6.1.
Another aspect of the connection between submodular functions and convexity is the submodular subdifferential. The polyhedra above can be defined for any (not necessarily submodular) set function. When the function is submodular however, it can be characterized efficiently. Firstly, note that for normalized submodular functions, for any $x \in \mathcal{P}_f$ the constraint at $Y = \emptyset$ is satisfied.

It is easy to see that the optimizers $s^*$ above form extreme points of the submodular polyhedron. Given a submodular function $f$ such that $f(\emptyset) = 0$, the condition that $x \in \mathcal{P}_f$ can be checked in polynomial time for every $x$. This follows directly from the fact that submodular function minimization is polynomial time.

**Proposition 2.2.** Given a submodular function $f$, checking if $x \in \mathcal{P}_f$ is equivalent to the condition $\min_{X \subseteq V}[f(X) - x(X)] \geq 0$, which can be checked in poly-time.

### 2.2 The Submodular Subdifferential

Another aspect of the connection between submodular functions and convexity is the submodular subdifferentials [16]. The subdifferential $\partial_f(X)$ of a submodular set function $f : 2^V \to \mathbb{R}$ for a set $Y \subseteq V$ is defined analogously to the subdifferential of a continuous convex function:

$$\partial_f(X) = \{ x \in \mathbb{R}^n : f(Y) - x(Y) \geq f(X) - x(X) \text{ for all } Y \subseteq V \} \quad (5)$$

The polyhedra above can be defined for any (not necessarily submodular) set function. When the function is submodular however, it can be characterized efficiently. Firstly, note that for normalized submodular functions, for any $h_X \in \partial_f(X)$, we have $f(X) - h_X(X) \leq 0$ which follows by the constraint at $Y = \emptyset$. Akin to the submodular polyhedron, the extreme points of the submodular subdifferential also admits interesting characterizations. We shall denote a subgradient at $X$ by $h_X \in \partial_f(X)$. Similar to the submodular polyhedron, the extreme points of $\partial_f(Y)$ may be computed via a greedy algorithm: Let $\sigma$ be a permutation of $V$ that assigns the elements in $X$ to the first $|X|$ positions ($i \leq |X|$ if and only if $\sigma(i) \in X$) and $S_{|X|} = X$. An illustration of this is shown in Figure 4.

![Figure 4](image-url)
This chain defines an extreme point $h^*_X$ of $\partial_f(X)$ with entries

$$h^*_X(\sigma(i)) = f(S^*_Y) - f(S^*_{Y-1}).$$

Note that for every subgradient $h_X \in \partial_f(X)$, we can define a modular lower bound $m_X(Y) = f(X) + h_X(Y) - h_X(X), \forall Y \subseteq V,$ which satisfies $m_X(Y) \leq f(Y), \forall Y \subseteq V$. Moreover, we have that $m_X(X) = f(X)$, and hence the subdifferential exactly correspond to the set of tight modular lower bounds of a submodular function, at a given set $X$. If we choose $h_X$ to be an extreme subgradient, the modular lower bound becomes $m_X(Y) = h_X(Y)$, resulting in a normalized modular function.

The subdifferential defined in Eqn. (6) is defined via an exponential number of inequalities. A key observation however is that many of these inequalities are redundant. Define three polyhedra:

$$\partial^1_f(X) = \{x \in \mathbb{R}^n : f(Y) - x(Y) \geq f(X) - x(X), \forall Y \subseteq X\}$$

$$\partial^2_f(X) = \{x \in \mathbb{R}^n : f(Y) - x(Y) \geq f(X) - x(X), \forall Y \supseteq X\}$$

$$\partial^3_f(X) = \{x \in \mathbb{R}^n : f(Y) - x(Y) \geq f(X) - x(X), \forall Y : Y \not\subseteq X, Y \not\supseteq X\}$$

We immediately have that $\partial_f(X) = \partial^1_f(X) \cap \partial^2_f(X) \cap \partial^3_f(X)$. The following Lemma shows that the inequalities in $\partial^3_f(X)$ are redundant in characterizing $\partial_f(X)$ when given $\partial^1_f(X)$ and $\partial^2_f(X)$.

**Lemma 2.3.** (\cite{1}, Lemma 6.4) Given a submodular function $f$, $\partial_f(X) = \partial^1_f(X) \cap \partial^2_f(X)$. Hence,

$$\partial_f(X) = \{\{x \in \mathbb{R}^n : f(Y) - x(Y) \geq f(X) - x(X) \text{ for all } Y \subseteq [0, X] \cup [X, V]\}$$

In the above, $[A, B] = \{X \subseteq V : A \subseteq X \subseteq B\}$ whenever $A \subseteq B$. We thus see that for $X \neq \{0, V\}$, many of the inequalities defining $\partial_f(X)$ are in fact redundant.

The subdifferential at the emptyset has a special relationship since $\partial_f(\emptyset) = P_f$. Similarly $\partial_f(V) = P_{f^\#}$, where $f^\#(X) = f(V) - f(V \setminus X)$ is the submodular dual of $f$. Furthermore, since $f^\#$ is a supermodular function, it holds that $\partial_f(V)$ is a supermodular polyhedron (for a supermodular function $g$, the supermodular polyhedron is defined as $P_g = \{x : x(X) \geq g(X), \forall X \subseteq V\}$).

The following lemma shows another instructive fact about the subdifferentials:

**Lemma 2.4.** (\cite{1}, Lemma 6.5) For any submodular function $f$, $\partial_f(X) = \partial_f(X) \times \partial_{f_X}(0)$, where $f^X(Y) = f(Y), \forall Y \subseteq X$, and $f_X(Y) = f(Y \cup X) - f(X), \forall Y \subseteq V \setminus X$, and $\times$ denotes the direct product.

Finally define:

$$\partial^{(1,1)}_f(X) = \{x \in \mathbb{R}^V : \forall j \in X, f(j|X \setminus j) \leq x(j) \text{ and } \forall j \not\in X, f(j|X) \geq x(j)\}.$$  

Notice that $\partial^{(1,1)}_f(X) \supseteq \partial_f(X)$ since we are reducing the constraints of the subdifferential. In particular $\partial^{(1,1)}_f(X)$ just considers $n$ inequalities, by choosing the sets $Y$ in Eqn. (10) such that $|Y \Delta X| = 1$ (i.e., Hamming distance one away from $X$). This polyhedron will be useful in characterizing local minimizers of a submodular function (see Section 4) and motivating analogous constructs for local maxima (see, for example, Proposition 8.2).

---

\(^3\)A set function $h$ is said to be normalized if $h(\emptyset) = 0$. 
2.3 Generalized Submodular Lower Polyhedron

In this section, we define a generalization of the submodular polyhedron, which we call the generalized submodular lower polyhedron. While this construct has not been defined explicitly before, we investigate it primarily with the aim of contrasting this with the results on the concave polyhedral aspects of a submodular function (Section 6).

Define the generalized submodular lower polyhedron as:

\[ P_{\text{gen}}^f = \{ (x,c) : x \in \mathbb{R}^n, c \in \mathbb{R} : x(X) + c \leq f(X), \forall X \subseteq V \} \]  

(12)

This generalized polyhedron \( P_{\text{gen}}^f \subseteq \mathbb{R}^{n+1} \) intuitively captures the affine (or unnormalized) modular lower bounds of \( f \). The definition above holds for any arbitrary set function, not necessarily submodular, in which case we call it the generalized lower polyhedron. In the case of submodular functions, this generalized lower polyhedron has interesting connections to the submodular polyhedron. In particular, note that \( P_{\text{gen}}^f \cap \{ (x,c) : c = 0 \} = \{ (x,c) : x \in P_f, c = 0 \} \). In other words, the slice \( c = 0 \) of the generalized submodular polyhedron is the submodular polyhedron of \( f \). Also notice that for a normalized submodular function \( f \), the constraint at \( X = \emptyset \), requires that \( c \leq 0 \).

The extreme points of \( P_{\text{gen}}^f \) also are easy to characterize when \( f \) is submodular. Surprisingly, all the extreme points lie exactly on the hyperplane \( c = 0 \) with \( x \) being the extreme points of \( P_f \).

**Lemma 2.5.** Given a set function \( f \), \( (x,c) \in P_{\text{gen}}^f \) lies on a face of the polyhedron if and only if there exists a set \( X \) such that \( x \in \partial_f(X) \) and \( c = f(X) - x(X) \).

**Proof.** Notice that \( (x,c) \) lies on a face of \( P_{\text{gen}}^f \) if and only if there exists a set \( X \) such that \( x(X) + c = f(X) \) and for all \( Y \subseteq V, x(Y) + c \leq f(Y) \). Since then \( x(Y) - x(X) \leq f(Y) - f(X) \), we have that \( x \in \partial_f(X) \) and \( c = f(X) - x(X) \) that, as mentioned above, has \( c \leq 0 \) when \( f \) is submodular. \( \square \)

The extreme points of \( P_{\text{gen}}^f \) also are easy to characterize when \( f \) is submodular. Surprisingly, all the extreme points lie exactly on the hyperplane \( c = 0 \) with \( x \) being the extreme points of \( P_f \).
Lemma 2.6. Given a submodular function $f$, $(x, c)$ is an extreme point of $P_f^{gen}$ if and only if $x$ is an extreme point of $P_f$ and $c = 0$. Furthermore, for any $y \in \mathbb{R}^n$,

$$\max_{(x,c) \in P_f^{gen}} [(x, y) + c] = \max \{ \max_{x \in \partial f(X)} [(x, y) + f(X) - x(X)] \mid X \subseteq V \} = \max_{x \in P_f} (x, y) \quad (13)$$

Proof. First we show that (i) $=$ (ii). Notice that a maximum in (i) (denoted by $(x^*, c^*)$) occurs at a face of $P_f^{gen}$ which, by Lemma 2.5, says there exists an $X$ such that $x^* \in \partial f(X)$ with $c^* = f(X) - x(X)$. This subdifferential is considered in (ii)’s outer max implying (ii) $\geq$ (i). Moreover, it is also easy to see that the point $(x, f(X) - x(X)) \in P_f^{gen}$, for any $x \in \partial f(X)$, from the definitions of the generalized submodular lower polyhedron and the subdifferential. Hence (ii) $\geq$ (i), since it is a max over a much larger set.

We then show that (i) $=$ (iii). It is immediate that (iii) $\leq$ (i) since (iii) is a more constrained case of (i) under $c = 0$. Next, we show (iii) $\geq$ (i), which states that for a submodular function, the linear program over the generalized submodular polyhedron is equivalent to a linear program over the submodular polyhedron. This result follows as a corollary from Lemma 2.1. Specifically, for any $(x, c) \in P_f^{gen}$, we have that

$$\max_{s \in P_f} w^T s = \sum_i \lambda_i f(S_i^{w c}) \geq \sum_i \lambda_i [(x, 1)S_i^{w c}] + c \geq (x, w) + c \quad (14)$$

where the last inequality follows from the facts that $\sum_i \lambda_i 1S_i^{w c} = w$ and $\sum_i \lambda_i = 1$. In particular, this also means that in the optimization problem in ii, the maximum over $X \subseteq V$ occurs at $X = \emptyset$, when $\partial f(X) = \partial f$.

Lastly, note that since every linear program over the generalized submodular polyhedron can be cast as a linear program over the submodular polyhedron, the extreme points of both polyhedra must also be the same. \hfill \square

Intuitively, $(x, c)$ is an extreme point if $x$ is an extreme point of a subdifferential $\partial f(X)$ for some set $X$. Since the extreme points of the subdifferentials are exactly the extreme points of the submodular polyhedron, the result follows.

Finally it is worth mentioning that similar to the submodular polyhedron, the generalized submodular polyhedron membership problem (i.e does $(x, c) \in P_f^{gen}$) is polynomial time, and can be solved via submodular minimization. This is again parallel to the submodular polyhedron.

Proposition 2.7. Given a submodular function $f$, $(x, c) \in P_f^{gen}$ if and only if $c \leq \min_{X \subseteq V} [f(X) - x(X)]$. Since submodular minimization is polynomial time, the generalized submodular polyhedral membership problem is also polynomial time.

A visualization of the generalized submodular lower polyhedron for a submodular function on $V = \{v_1, v_2\}$ is shown in Figure 5.

3 Convex extensions of a Submodular Function

We now introduce the convex extension of a submodular functions. We shall see a number of equivalent ways to characterize this extension and observe how they can be computed very efficiently as what is known as the Lovász extension \[39\]. The results of this section are mainly from \[39\] \[8\] \[49\]. Following \[49\] \[8\], we consider two main characterizations of the convex extensions, as what we call
polyhedral characterization and distributional characterization. Again, the main purpose of this section, is to review the existing work, and contrast the results with the ones we shall see in Section 7 on the concave extensions of submodular functions.

3.1 Polyhedral characterization of the convex extensions

The convex extension of any set function (not necessarily submodular) can be seen as the pointwise supremum of convex functions which lower bound the set function \( f \). More clearly, let

\[
\Phi_f = \{ \phi : \phi \text{ is convex in } [0, 1]^V \text{ and } \phi(1_X) \leq f(X), \forall X \subseteq V \}.
\]

(15)

Then define:

\[
\hat{f}(w) = \max_{\phi \in \Phi_f} \phi(w), \text{ for } w \in [0, 1]^n
\]

(16)

It is not hard to show that \( \hat{f} \) is convex and satisfies the relation \( \hat{f}(1_X) = f(X) \). The above expression can in fact be simplified for any set function, and it suffices to consider affine lower instead of convex lower bounds. In particular Eqn. (16) can be expressed as a linear program over the generalized polyhedron.

**Lemma 3.1.** Given a set function \( f \), the convex extension of \( f \) in Eqn. (16) can be expressed as:

\[
\hat{f}(w) = \max_{(x,c) \in P_{gen} f} [\langle x,w \rangle + c], \forall w \in [0, 1]^n
\]

(17)

**Proof.** The proof of the equivalence follows from a simple observation. For a given \( w \), let \( \hat{\phi} \) be an argmax in Eqn. (16). Then since \( \hat{\phi} \) is a convex function in \([0, 1]^V\), there exists a subgradient \( x \in \mathbb{R}^n \) at \( w \) and value \( d \), such that \( \langle x,y \rangle + d \leq \hat{\phi}(y), \forall y \) and \( \langle x,w \rangle + d = \hat{\phi}(w) \). In other words, \( \langle x,y \rangle + d \) is a linear lower bound of \( \hat{\phi}(y) \), tight at \( w \). Hence, at value \( w \), \( \hat{f}(w) \) takes value \( \langle x,w \rangle + d \). Finally notice that \( (x,d) \in P_{gen} f \) since \( x(X) + d \leq \hat{\phi}(1_X) \leq f(X), \forall X \subseteq V \).

So far, in the above, we have not invoked the submodularity of \( f \). If \( f \) is submodular however, the above polyhedral characterization can be replaced by an linear program over the submodular polyhedron. In other words,

**Lemma 3.2.** For a submodular function \( f \), the expressions in Eqn. (16), (17) can be rewritten as:

\[
\hat{f}(w) = \max_{s \in P_f} w^T s, \forall w \in [0, 1]^n.
\]

(18)

**Proof.** This follows directly from Lemma 2.6.

The above result is not surprising given that the extreme points of \( P_{gen} f \) are identical to the extreme points of \( P_f \), when \( f \) is submodular.

3.2 Distributional characterization of the convex extension

Another way to characterize the continuous extension of a set function \( f \) is as follows. Denote \( \Lambda_w \) as the set:

\[
\Lambda_w = \left\{ \{\lambda_S, S \subseteq V\} : \sum_{S \subseteq V} \lambda_S 1_S = w, \sum_{S \subseteq V} \lambda_S = 1, \text{ and } \forall S, \lambda_S \geq 0 \right\}, \text{ for } w \in [0, 1]^n
\]

(19)
Then the convex extension \( \hat{f} \) can be equivalently written as:

\[
\hat{f}(w) = \min_{\lambda \in \Lambda_w} \sum_{S \subseteq V} \lambda_S f(S) \quad (20)
\]

The reason this representation is called distributional is that the convex extension here is computed by minimizing over particular distributions over sets. Again, it is not hard to see that this characterization is a convex extension.

For a submodular function, the distribution characterization takes on a nice form, which is known classically as the Lovász extension. This result can be found, for example, in [8, 49]:

**Lemma 3.3.** [8, 39, 9] Given a submodular function \( f \),

\[
\hat{f}(w) = \sum_{i=1}^{n} w(\sigma_w(i)) (f(S_{i}^{\sigma_w}) - f(S_{i-1}^{\sigma_w})) + \sum_{i=1}^{n-1} (w(\sigma_w(i)) - w(\sigma_w(i + 1))) f(S_{i}^{\sigma_w}),
\]

where \( \sigma_w \) is a permutation satisfying \( w(\sigma_w(1)) \geq w(\sigma_w(2)) \geq \cdots \geq w(\sigma_w(n)). \)

It is clear from above that the minimizing distribution \( \lambda \) is a form of a chain distribution, where the chain here is the sequence of sets \( S_0^{\sigma_w}, S_1^{\sigma_w}, \cdots, S_n^{\sigma_w} \) defined in Lemma 2.1. We also see the relationship between the two characterizations in the case of submodular functions, since Eqn. (21) is exactly the solution of the linear program over the submodular polyhedron (see Lemma 2.1). Hence the two forms of convex extensions, i.e. the distributional characterization from Lemma 3.3 and polyhedral characterization from Lemma 3.2 are identical for a submodular function. The resulting convex function \( \hat{f} \) is the Lovász extension.

The equivalence between the two characterizations holds for general set functions, not necessarily submodular. In other words, Eqn. (20) and Eqn. (16), (17) are identical for any set function. This follows directly from the arguments in [8, 49]. The only catch, however, is that Lemmas 3.2 and 3.3 do not hold for general set functions and \( \hat{f} \) can be NP hard to evaluate in general [8, 49, 3].

### 3.3 Convex Extensions and Submodular Minimization

The Lovász extension plays an important role in submodular minimization. In particular, minimizing the Lovász extension is equivalent to minimizing a submodular function:

**Lemma 3.4.** [39] Given a submodular function \( f \),

\[
\min_{X \subseteq V} f(X) = \min_{x \in [0,1]^n} \hat{f}(x)
\]

Furthermore, given the minimizer \( x^* \) of the RHS above, we can obtain a set \( X^* \) such that \( f(X^*) = \hat{f}(x^*) \).

This implies that unconstrained submodular minimization has an integrality gap of one and the two problems are equivalent.

### 4 Optimality conditions for submodular minimization

Fujishige [18] provides some interesting characterizations to the optimality conditions for unconstrained submodular minimization. The following theorem can be thought of as a discrete analog to the KKT conditions:
Lemma 4.1. ([17], Lemma 7.1) A set $A \subseteq V$ is a minimizer of $f : 2^V \rightarrow \mathbb{R}$ if and only if:

$$0 \in \partial f(A)$$

(23)

This immediately provides necessary and sufficient conditions for optimality of $f$:

Lemma 4.2. ([19], Theorem 7.2) A set $A$ minimizes a submodular function $f$ if and only if

$$f(A) \leq f(B)$$

for all sets $B$ such that $B \subseteq A$ or $A \subseteq B$.

In other words, it is sufficient to check only the subsets and supersets of $A$ to ensure that $A$ is a global optimizer of $f$. The above Lemma follows from Eqn. 10 and Lemma 4.1. Analogous characterizations have also been provided for constrained forms of submodular minimization, and interested readers may look at [18]. Finally, we can provide a simple characterization on the local minimizers of a submodular function.

Lemma 4.3. A set $A \subseteq V$ is a local minimizer of a submodular function if and only if $0 \in \partial f(\triangle(A))$.

As was shown in [30], a local minimizer of a submodular function, in the unconstrained setting, can be found efficiently in $O(n^2)$ complexity.

While unconstrained submodular minimization is easy, most forms of constrained submodular minimization become NP hard. For example, a simple cardinality lower bound constraint makes the problem of submodular minimization (even with monotone submodular functions) NP hard without even constant factor approximation guarantees [48]. These results, however, can be extended when the constraints are lattice constraints [19] in which case many of the results above still hold.

5 Convex Characterizations: Discrete Separation Theorem and Fenchel Duality Theorem

Finally we review some interesting theorems which characterize convex functions, but also go through for submodular functions.

5.1 The Discrete Separation Theorem (DST)

The discrete separation theorem known in context of convexity, states that given a convex function $\phi$ and a concave function $\psi$ such that $\forall x, \phi(x) \geq \psi(x)$, there exists an affine function $\langle h, x \rangle + c$ such that $\forall x, \psi(x) \geq \langle h, x \rangle + c \geq \psi(x)$. A similar relation holds for submodular functions.

The lemma below was shown by Frank [15] in the context of submodular functions:

Lemma 5.1. [15], [19, Theorem 4.12] Given a submodular function $f$ and a supermodular function $g$ such that $f(X) \geq g(X)$, $\forall X$ (and which satisfy $f(\emptyset) = g(\emptyset) = 0$), there exists a modular function $h$ such that $f(X) \geq h(X) \geq g(X)$. Furthermore, if $f$ and $g$ are integral so may be $h$.

This Lemma can also be shown through the Lovász extension. In particular, given a submodular function $f$ and a supermodular function $g$ such that $f(X) \geq g(X)$, $\forall X$, we can construct the convex and concave extensions $\hat{f}$ and $\hat{g}$ of $f$ and $g$ (the concave extension $\hat{g}$ can be constructed via the Lovász extension of $-g$). From the expressions of $\hat{f}$ and $\hat{g}$, it is not hard to see that $\hat{f}(x) \geq \hat{g}(x), \forall x$. Hence using the Discrete Separation Theorem from convex analysis, we can find a linear function $h$, which when restricted to 0/1 vectors, gives the modular function $h$.

$^4$A set $A$ is a local minimizer of a submodular function if $f(X) \geq f(A), \forall X : |X\setminus A| \leq 1$, and $|A\setminus X| = 1$, that is all sets $X$ no more than hamming-distance one away from $A$. 


5.2 Fenchel Duality Theorem (FDT)

The Fenchel duality theorem in the context of convexity [44] provides a relation between the minimizers of the function and it is dual. Given a convex function $\phi$ and a concave function $\psi$, the Fenchel dual $\phi^*$ of $\phi$, and $\psi^*$ of $\psi$ is given as follows:

$$
\phi^*(y) = \max_{x \in \text{dom}(\phi)} [(x,y) - \phi(x)], \quad \psi^*(y) = \min_{x \in \text{dom}(\psi)} [(x,y) - \psi(y)].
$$

(24)

The dual functions $\phi^*$ and $\psi^*$ are convex and concave respectively. The Fenchel duality theorem then states that:

$$
\min_x [\phi(x) - \psi(x)] = \max_y [\psi^*(y) - \phi^*(y)]
$$

(25)

Analogous characterizations also hold for submodular functions [18]. Given a submodular function $f$ (or equivalently supermodular function $g$), the Fenchel dual $f^*$ (or equivalently $g^*$) is:

$$
f^*(x) = \max_{X \subseteq V} [x(X) - f(X)], \quad g^*(x) = \min_{X \subseteq V} [x(X) - g(X)].
$$

(26)

The Fenchel duals $f^*$ and $g^*$ are convex and concave functions respectively. Then the following Lemma holds:

**Lemma 5.2.** ([19, Theorem 6.3]) Given a submodular function $f$ and a supermodular function $g$, it holds that:

$$
\min_{X \subseteq V} [f(X) - g(X)] = \max_x [g^*(x) - f^*(x)].
$$

(27)

Further if $f$ and $g$ are integral, the maximum on the right hand side is attained by an integral vector $x$.

5.3 The Minkowski-Sum theorem

Submodular polyhedra and also the subdifferentials have an interesting characterization related to Minkowski sums.

**Lemma 5.3.** ([19, Theorem 6.8]) Given two submodular functions $f_1$ and $f_2$, it holds that that (the addition of the polyhedra below corresponds to a point-wise addition):

$$
\mathcal{P}_{f_1+f_2} = \mathcal{P}_{f_1} + \mathcal{P}_{f_2}, \quad \partial_{f_1+f_2}(X) = \partial_{f_1}(X) + \partial_{f_2}(X)
$$

(28)

Similarly it holds that, $\mathcal{P}_{f_1+f_2}^{\text{gen}} = \mathcal{P}_{f_1}^{\text{gen}} + \mathcal{P}_{f_2}^{\text{gen}}$.

The Minkowski Sum Theorem for the generalized submodular polyhedron, follows directly from the definition.

6 Concave Polyhedral Aspects of Submodular Functions

In this section, we investigate several polyhedral aspects of submodular functions relating them to concavity, thus complementing the results from Section 2. This provides a complete picture on the relationship between submodularity, convexity and concavity. We investigate the submodular upper polyhedron, submodular superdifferential and the generalized submodular upper polyhedron.
Figure 6: The submodular upper Polyhedron \( \mathcal{P}^f \) in two dimensions.

6.1 The submodular upper polyhedron

A first step in characterizing the concave aspects of a submodular function is the submodular upper polyhedron. Intuitively this is the set of tight modular upper bounds of the function, and we define it as follows:

\[
\mathcal{P}^f = \{ x \in \mathbb{R}^n : x(S) \geq f(S), \forall S \subseteq V \}
\]  

The above polyhedron can in fact be defined for any set function. In particular, when \( f \) is supermodular, we get what is known as the supermodular polyhedron [19]. Presently, we are interested in the case when \( f \) is submodular and hence we call this the submodular upper polyhedron. Interestingly this has a very simple characterization.

Lemma 6.1. Given a submodular function \( f \),

\[
\mathcal{P}^f = \{ x \in \mathbb{R}^n : x(j) \geq f(j) \}
\]

Proof. Given \( x \in \mathcal{P}^f \) and a set \( S \), we have \( x(S) = \sum_{i \in S} x(i) \geq \sum_{i \in S} f(i) \), since \( \forall i, x(i) \geq f(i) \) by Eqn. (29). Hence \( x(S) \geq \sum_{i \in S} f(i) \geq f(S) \). Thus, the irredundant inequalities are the singletons. \( \square \)

Thus, the submodular upper polyhedron has a particularly simple characterization due to the submodularity of \( f \). In other words, this polyhedron is not polyhedrally tight in that many of the defining inequalities are redundant. This polyhedron alone is not particularly interesting to define a concave extension. We shall define a generalization of this in Section 6.3 that shall prove useful in characterizing the concave extension.

We end this section by investigating the submodular upper polyhedron membership problem. Owing to its simplicity, this problem is particularly simple which might seem surprising at first glance since \( x \in \mathcal{P}^f \) is equivalent to, Eqn. (29), checking if \( \max_{X \subseteq V} f(X) - x(X) \leq 0 \). This involves maximization of a submodular function which is NP hard. The following lemma shows that this problem is actually easy.

Corollary 6.2. Given a submodular function \( f \) and vector \( x \), let \( X \) be a set such that \( f(X) - x(X) > 0 \). Then there exists an \( i \in X : f(i) - x(i) > 0 \).

Proof. Observe that \( f(X) - x(X) \leq \sum_{i \in X} f(i) - x(i) \). Since the l.h.s. is greater than 0, it implies that \( \sum_{i \in X} f(i) - x(i) > 0 \). Hence there should exist an \( i \in X \) such that \( f(i) - x(i) > 0 \). \( \square \)

Thus, it is sufficient to check the singleton values, i.e \( f(i) - x(i) \), and if all these are less than or equal to zero, then \( x \in \mathcal{P}^f \). This also follows immediately from Lemma 6.1.

An interesting corollary of the above, is that it is in fact easy to check if the maximizer of a submodular function is greater than equal to zero. Given a submodular function \( f \), the problem is
whether \( \max_{X \subseteq V} f(X) \geq 0 \). This can easily be checked without resorting to submodular function maximization.

**Corollary 6.3.** Given a submodular function \( f \) with \( f(\emptyset) = 0 \), \( \max_{X \subseteq V} f(X) > 0 \) if and only if there exists an \( i \in V \) such that \( f(i) > 0 \).

**Proof.** If for any \( j \), \( f(j) > 0 \) it implies that \( \max_{X \subseteq V} f(X) \geq f(j) > 0 \). On the other hand, if \( \forall j, f(j) \leq 0 \), we have that \( \forall X \subseteq V, f(X) \leq \sum_{i \in X} f(i) \leq 0 \). Hence \( \max_{X \subseteq V} f(X) = 0 \).

This fact is true only for a submodular function. For general set functions, even when \( f(\emptyset) = 0 \), it could potentially require an exponential search to determine if \( \max_{X \subseteq V} f(X) > 0 \).

### 6.2 The Submodular Superdifferentials

Given a submodular function \( f \) we can characterize its superdifferential as follows. We denote for a set \( X \), the superdifferential with respect to \( X \) as \( \partial_f(X) \). Observe then that,

\[
\partial_f(X) = \{ x \in \mathbb{R}^n : f(Y) - x(Y) \leq f(X) - x(X), \forall Y \subseteq X \} \tag{31}
\]

This characterization is analogous to the subdifferential of a submodular function defined in Eqn. [5]. This is also akin to the superdifferential corresponding to a continuous concave function.

Each supergradient \( g_X \in \partial_f(X) \), defines a modular upper bound of a submodular function. In particular, define \( m^X(Y) = g_X(Y) + f(X) - g_X(X) \). Then \( m^X(Y) \) is a modular function which satisfies \( m^X(Y) \leq f(Y), \forall Y \subseteq X \) and \( m^X(X) = f(X) \).

We note that \( (x(v_1), x(v_2), \ldots, x(v_n)) = (f(v_1), f(v_2), \ldots, f(v_n)) \in \partial_f(\emptyset) \) which shows that at least \( \partial_f(\emptyset) \) exists. A bit further below (specifically Theorem 6.8) we show that for any submodular function, \( \partial_f(X) \) is non-empty for all \( X \subseteq V \).

Note that the superdifferential is defined by an exponential (i.e., \( 2^{|V|} \)) number of inequalities. However owing to the submodularity of \( f \) and akin to the subdifferential of \( f \), we can reduce the number of inequalities. Define three polyhedrons as:

\[
\partial_1^f(X) = \{ x \in \mathbb{R}^n : f(Y) - x(Y) \leq f(X) - x(X), \forall Y \subseteq X \} \tag{32}
\]
\[
\partial_2^f(X) = \{ x \in \mathbb{R}^n : f(Y) - x(Y) \leq f(X) - x(X), \forall Y \supseteq X \} \tag{33}
\]
\[
\partial_3^f(X) = \{ x \in \mathbb{R}^n : f(Y) - x(Y) \leq f(X) - x(X), \forall Y : Y \not\subseteq X, Y \not\supseteq X \} \tag{34}
\]

A trivial observation is that: \( \partial_f(X) = \partial_1^f(X) \cap \partial_2^f(X) \cap \partial_3^f(X) \). As we show below for a submodular function \( f \), \( \partial_1^f(X) \) and \( \partial_3^f(X) \) are actually very simple polyhedra.

**Lemma 6.4.** For a submodular function \( f \),

\[
\partial_1^f(X) = \{ x \in \mathbb{R}^n : f(j \mid X \setminus j) \geq x(j), \forall j \in X \} \tag{35}
\]
\[
\partial_3^f(X) = \{ x \in \mathbb{R}^n : f(j \mid X \setminus j) \leq x(j), \forall j \not\in X \}. \tag{36}
\]

**Proof.** Consider \( \partial_1^f(X) \). Notice that the inequalities defining the polyhedron can be rewritten as \( \partial_1^f(X) = \{ x \in \mathbb{R}^n : x(X \setminus Y) \leq f(X) - f(Y), \forall Y \subseteq X \} \). We then have that \( x(X \setminus Y) = \sum_{j \in X \setminus Y} x(j) \leq \sum_{j \in X \setminus Y} f(j \mid X \setminus j) \), since \( \forall j \in X, x(j) \leq f(j \mid X \setminus j) \) (this follows by considering only the subset of inequalities of \( \partial_1^f(X) \) with sets \( Y \) such that \( \mid X \setminus Y \mid = 1 \)). Hence \( x(X \setminus Y) \leq \sum_{j \in X \setminus Y} f(j \mid X \setminus j) \leq f(X) - f(Y) \). Hence an irredundant set of inequalities include those defined only through the singletons.
In order to show the characterization for \( \partial f^1(X) \), we have that \( \partial f^1(X) = \{ x \in \mathbb{R}^n : (x(Y \setminus X) \geq f(Y) - f(X), \forall Y \subseteq X \} \). It then follows that, \( x(Y \setminus X) = \sum_{j \in Y \setminus X} x(j) \geq \sum_{j \in Y \setminus X} f(j|x), \) since \( \forall j \notin X, x(j) \geq f(j|X) \). Hence \( x(X \setminus Y) = \sum_{j \notin Y \setminus X} f(j|X) \geq (Y) - f(X), \) and again, an irredundant set of inequalities include those defined only through the singletons.

The above characterization significantly reduces the inequalities governing \( \partial f^1(X) \), and in fact, the polytopes \( \partial f^1(X) \) and \( \partial f^2(X) \) are very simple polyhedra. Recall that this is analogous to the submodular subdifferential, where again owing to submodularity the number of inequalities are reduced significantly. In that case, we just need to consider the sets \( Y \) which are subsets and supersets of \( X \). It is interesting to note the contrast between the redundancy of inequalities in the subdifferentials and the superdifferentials. In particular, here, the inequalities corresponding to sets \( Y \) being the subsets and supersets of \( X \) are mostly redundant, while the non-redundant ones are the rest of the inequalities. In other words, in the case of the subdifferential, \( \partial f^1(X) \) and \( \partial f^2(X) \) were non-redundant, while \( \partial f^1(X) \) was entirely redundant given the first two. In the case of the superdifferentials, \( \partial f^1(X) \) and \( \partial f^2(X) \) are mostly internally redundant (they can be represented using only by \( n \) inequalities), while \( \partial f^1(X) \) has no redundancy in general.

In order to gain more intuition for the superdifferentials, we consider some examples in both two and three dimensions.

**Example 6.1.** See Figure 7. We consider here superdifferentials when \( V = \{1, 2\} \). Then from the lemma above, \( \partial f^1(\emptyset) = \{ x \in \mathbb{R}^2 : f(1|\emptyset) \leq f(j), \forall j \in \{1, 2\} \} \). Similarly \( \partial f^1(\{1, 2\}) = \{ x \in \mathbb{R}^2 : f(\emptyset|\{1, 2\}) \geq f(j|\{1, 2\}), \forall j = \{1, 2\} \} \). Now consider \( \partial f^1(1) \). Then the governing inequalities for this are:

\[
\partial f^1(1) = \{ x \in \mathbb{R}^2 : x_1 \leq f(\{1\}), x_2 \geq f(\{2\}) \geq 1, x_1 - x_2 \leq f(\{1\}) - f(\{2\}) \}
\]

The extreme points of this polyhedron are the vectors \( \{f(\{1\}), f(\{2\})\} \) and \( \{f(\{1\}), f(\{2\})\} \). The way we obtain the extreme points is as follows. Setting the inequalities \( 37 \) and \( 39 \) as equalities, we get the extreme point \( \{f(\{1\}), f(\{2\})\} \). The inequality \( 38 \), then is \( x_2 = f(\{2\}) \geq f(\{2\}) \geq 1 \), which holds. We then set inequalities \( 38 \) and \( 39 \) as equalities, which gives \( x_1 = f(\{1\}) \geq f(\{2\}) \) and \( x_1 = f(\{1\}) \geq f(\{2\}) \), thus giving the second extreme point \( \{f(\{1\}), f(\{2\})\} \). It is easy to see that
inequality (37) is satisfied. Finally, if we set inequalities (37) and (38) as equalities, we get
\[ x_1 = f(\{1\}), \quad x_2 = f(\{2\} | \{1\}). \]
Then we have that
\[ x_1 - x_2 = f(\{1\}) - f(\{2\} | \{1\}) = 2f(\{1\}) - f(\{1\}, \{2\}). \]
Inequality (39), then requires,
\[ 2f(\{1\}) - f(\{1, 2\}) \geq f(\{1\}) - f(\{2\}) \Rightarrow f(\{1\}) + f(\{2\}) \leq f(\{1, 2\}), \]
which does not hold. Hence the only extreme points are the two vectors above. One can similarly
investigate \( \partial f(\{2\}) \), which has the same extreme points.

It is clear from the above example that superdifferentials in the two-dimensional case are easy
to find and characterize. However this is not the case in three dimensions where the shape of the
superdifferentials depends strongly on the particular submodular functions.

**Example 6.2.** Let \( V = \{1, 2, 3\} \). Recall that \( f(\emptyset) = 0 \). Then consider
\( \partial f(\{1\}) = \{ x \in \mathbb{R}^3 : f(\emptyset) \leq f(\emptyset) \} \). This polyhedron can be represented via the following
irredundant inequalities:

\[
\partial f(\{1\}) = \{ x \in \mathbb{R}^3 : x_1 \leq f(\emptyset), \quad \text{for } Y = \emptyset \}
\]
\[
\quad x_2 \geq f(\{2\} | \{1\}), \quad \text{for } Y = \{1, 2\} \tag{40}
\]
\[
\quad x_2 - x_1 \geq f(\{2\}) - f(\{1\}), \quad \text{for } Y = \{2\} \tag{41}
\]
\[
\quad x_3 \geq f(\{3\} | \{1\}), \quad \text{for } Y = \{1, 3\} \tag{42}
\]
\[
\quad x_3 - x_1 \geq f(\{3\}) - f(\{1\}), \quad \text{for } Y = \{3\} \tag{43}
\]
\[
\quad x_2 + x_3 - x_1 \geq f(\{2, 3\}) - f(\{1\}), \quad \text{for } Y = \{2, 3\} \tag{44}
\]

The other two inequalities (for \( Y = \{1, 2, 3\} \) and \( Y = \{1\} \)) are redundant. We now consider
the extreme points of this polyhedron. We consider Eqns. (40), (42), (44) with equality, and we
obtain an extreme point \( \{ f(\{1\}), f(\{2\}), f(\{3\}) \} \). It is easy to see that all other inequalities are
satisfied. Consider next Eqns. (42), (44), and (45) with equality and we get a potential extreme
point \( \{ f(\{1\}) + f(\{2, 3\}) - f(\{2\}) - f(\{3\}), f(\{2\}) + f(\{3\}) | \{2\} \} \). Observe that
\( x_1 = f(\{1\}) + f(\{2, 3\}) - f(\{2\}) - f(\{3\}) \leq f(\{1\}) \), and hence Eqn. (40) is satisfied. However \( x_2 = f(\{2\}) \)
may be bigger or smaller than \( f(\{2\} | \{1\}) \) and Eqn. (41) may or may not be violated. Similarly,
\( x_3 = f(\{3\} | \{2\}) \) is not comparable to \( f(\{3\} | \{1\}) \), and hence Eqns. (43) might or might not be violated.
Consequently we cannot determine if by combining together Eqns. (42), (44), and (45), we obtain an
extreme point, unless we have more information about the current submodular function being used.
Hence, we see from this example that we cannot hope to find the extreme points both analytically and
generically.

The above example shows that a particular expression (obtained via a combination of inequalities)
might or might not be extreme, depending on the particular submodular function and its valuation.
This is unlike the subdifferential, where an analytical expression is always extreme for all
submodular functions. Thus, unlike the subdifferentials, we cannot expect a closed form expression for
the extreme points of \( \partial f(Y) \). Moreover, they also seem to be hard to characterize algorithmically. For
example, the superdifferential membership problem is NP hard.

**Lemma 6.5.** Given a submodular function \( f \) and a set \( Y : \emptyset \subset Y \subset V \), the membership problem
\( y \in \partial f(Y) \) is NP hard.

**Proof.** Notice that the membership problem \( y \in \partial f(Y) \) is equivalent to asking
\( \max_{X \subseteq V} f(X) - y(X) \leq f(Y) - y(Y) \). In other words, this is equivalent to asking if \( Y \) is a maximizer of \( f(X) - y(X) \) for a given
vector \( y \). This is the decision version of the submodular maximization problem and correspondingly
is NP hard when \( \emptyset \subset Y \subset V \). \( \square \)
We can then define the outer bound:

\[ \partial^f \Delta(k,l)(X) = \partial^f_1(X) \cap \partial^f_2(X) \cap \partial^f_3(X) \cap \partial^f_4(X). \]  

Observe that \( \partial^f_\Delta(k,l)(X) \) is expressed in terms of \( O(n^{k+l}) \) inequalities, and hence for a given \( k, l \) we can obtain the representation of \( \partial^f_\Delta(k,l)(X) \) in polynomial time. We will see that this provides us with a hierarchy of outer bounds on the superdifferential:

**Theorem 6.7.** For a submodular function \( f \):

1. \( \partial^f_\Delta(1,1)(X) = \partial^f_1(X) \cap \partial^f_2(X) \)

2. \( \forall 1 \leq k' \leq k, 1 \leq l' \leq l, \partial^f(X) \subseteq \partial^f_\Delta(k,l)(X) \subseteq \partial^f_\Delta(k',l')(X) \subseteq \partial^f_\Delta(1,1)(X) \)

3. \( \partial^f_\Delta(n,n)(X) = \partial^f(X). \)

**Proof.** The proofs of items 1 and 3 follow directly from definitions. To see item 2, notice that the polyhedra \( \partial^f_\Delta(k,l) \) become tighter as \( k \) and \( l \) increase finally approaching the superdifferential. \( \square \)

Similar to the submodular subdifferential, we shall call \( \partial^f_\Delta(1,1)(Y) \) as the local approximation of the superdifferential. In particular,

\[ \partial^f_\Delta(1,1)(X) = \{ x \in \mathbb{R}^n : x(j) \leq f(j|X \setminus j) \forall j \in X, x(j) \geq f(j|X) \forall j \notin X \}. \]  

We shall see in Section 8 that these outer bounds have interesting connections with approximation algorithms for submodular maximization.
6.2.2 Inner Bounds on the Superdifferential

While it is hard to characterize the extreme points of the superdifferential, we can provide some specific supergradients. Define three vectors as follows:

\[
\hat{g}_X(j) = \begin{cases} 
  f(j|X - j) & \text{if } j \in X \\
  f(j) & \text{if } j \notin X 
\end{cases}
\]  
(49)

\[
\tilde{g}_X(j) = \begin{cases} 
  f(j|V - j) & \text{if } j \in X \\
  f(j|X) & \text{if } j \notin X 
\end{cases}
\]  
(50)

\[
\bar{g}_X(j) = \begin{cases} 
  f(j|V - j) & \text{if } j \in X \\
  f(j) & \text{if } j \notin X 
\end{cases}
\]  
(51)

Then we have the following theorem:

**Theorem 6.8.** For a submodular function \( f \), \( \hat{g}_X, \tilde{g}_X, \bar{g}_X \in \partial f(X) \). Hence for every submodular function \( f \) and set \( X \), \( \partial f(X) \) is non-empty.

**Proof.** For submodular \( f \), the following bounds are known to hold [43]:

\[
f(Y) \leq f(X) - \sum_{j \in X \setminus Y}^{} f(j|X \setminus j) + \sum_{j \in Y \setminus X}^{} f(j|X \cap Y),
\]  
(52)

\[
f(Y) \leq f(X) - \sum_{j \in X \setminus Y}^{} f(j|X \cup Y \setminus j) + \sum_{j \in Y \setminus X}^{} f(j|X),
\]  
(53)

Using submodularity, we can loosen these bounds further to provide tight modular upper bounds [1, 33, 32, 26, 25]:

\[
f(Y) \leq f(X) - \sum_{j \in X \setminus Y}^{} f(j|X - \{j\}) + \sum_{j \in Y \setminus X}^{} f(j|\emptyset),
\]  
(54)

\[
f(Y) \leq f(X) - \sum_{j \in X \setminus Y}^{} f(j|V - \{j\}) + \sum_{j \in Y \setminus X}^{} f(j|X),
\]  
(55)

\[
f(Y) \leq f(X) - \sum_{j \in X \setminus Y}^{} f(j|V - \{j\}) + \sum_{j \in Y \setminus X}^{} f(j|\emptyset).
\]  
(56)

From the three bounds above, and substituting the expressions of the supergradients, we may immediately verify that these are supergradients, namely that \( \hat{g}_X, \tilde{g}_X, \bar{g}_X \in \partial f(X) \). For example, starting with Eqn. (54), we have:

\[
f(Y) \leq f(X) - \sum_{j \in X \setminus Y}^{} f(j|X - \{j\}) + \sum_{j \in Y \setminus X}^{} f(j|\emptyset)
\]  
(57)

\[
= f(X) - \sum_{j \in X}^{} f(j|X \setminus \{j\}) + \sum_{j \in X \setminus Y}^{} f(j|X \setminus \{j\}) + \sum_{j \in Y \setminus X}^{} f(j|\emptyset)
\]  
(58)

\[
= f(X) - \hat{g}_X(X) + \tilde{g}_X(Y)
\]  
(59)

\]
These supergradients characterize inner bounds for the superdifferential. Define two polyhedra:

\[
\partial_{i}^{f}(X) = \{ x \in \mathbb{R}^{n} : f(j) \leq x(j), \forall j \notin X \},
\]

\[
\partial_{i}^{f}(X) = \{ x \in \mathbb{R}^{n} : f(j|V\setminus j) \geq x(j), \forall j \in X \}.
\]

Then define:

\[
\partial_{i,1}^{f}(X) = \partial_{i}^{f}(X) \cap \partial_{i}^{f}(X) = \{ x \in \mathbb{R}^{n} : f(j|X\setminus j) \geq x(j), \forall j \in X \text{ and } f(j) \leq x(j), \forall j \notin X \}
\]

\[
\partial_{i,2}^{f}(Y) = \partial_{i}^{f}(Y) \cap \partial_{i}^{f}(Y) = \{ x \in \mathbb{R}^{n} : f(j|V\setminus j) \geq x(j), \forall j \in X \text{ and } f(j|X) \leq x(j), \forall j \notin X \}
\]

\[
\partial_{i,3}^{f}(Y) = \partial_{i}^{f}(Y) \cap \partial_{i}^{f}(Y) = \{ x \in \mathbb{R}^{n} : f(j|V\setminus j) \geq x(j), \forall j \in X \text{ and } f(j) \leq x(j), \forall j \notin X \}.
\]

Then note that \(\partial_{i,1}^{f}(Y)\) is a polyhedron with \(\tilde{g}_{Y}\) as an extreme point. Similarly \(\partial_{i,2}^{f}(Y)\) has \(\tilde{g}_{Y}\), while \(\partial_{i,3}^{f}(Y)\) has \(\tilde{g}_{Y}\) as its extreme points. All these are simple polyhedra, with a single extreme point. Also define,

\[
\partial_{i,(1,2)}^{f}(Y) = \text{conv}(\partial_{i,1}^{f}(Y), \partial_{i,2}^{f}(Y))
\]

where \(\text{conv}(., .)\) represents the convex combination of two polyhedra.\(^5\) Then \(\partial_{i,(1,2)}^{f}(Y)\) is a polyhedron which has \(\tilde{g}_{Y}\) and \(\tilde{g}_{Y}\) as its extreme points. The following lemma characterizes the inner bounds of the superdifferential:

**Lemma 6.9.** Given a submodular function \(f\),

\[
\partial_{i,3}^{f}(Y) \subseteq \partial_{i,2}^{f}(Y) \subseteq \partial_{i,(1,2)}^{f}(Y) \subseteq \partial^{f}(Y),
\]

\[
\partial_{i,2}^{f}(Y) \subseteq \partial_{i,1}^{f}(Y) \subseteq \partial_{i,(1,2)}^{f}(Y) \subseteq \partial^{f}(Y)
\]

**Proof.** The proof of this lemma follows directly from the definitions of the supergradients, corresponding polyhedra and submodularity.

\(\square\)

### 6.2.3 Connections between the subdifferential and superdifferential at \(X\)

Finally we point out interesting connections between \(\partial_{f}(X)\) and \(\partial^{f}(X)\). Firstly, it is clear from the definitions that \(\partial_{f}(X) \subseteq \partial^{f,(1,1)}(X)\) and \(\partial^{f}(X) \subseteq \partial^{f,(1,1)}(X)\). Notice also that both \(\partial_{f}^{f,(1,1)}(X)\) and \(\partial^{f,(1,1)}(X)\) (from eqns (11) and (48) respectively) are simple polyhedra with a single extreme point,

\[
\tilde{g}_{X}(j) = \begin{cases} f(j|X - j) & \text{if } j \in X \\ f(j|X) & \text{if } j \notin X \end{cases}
\]

The point \(\tilde{g}_{X}\), is in general, neither a subgradient nor a supergradient at \(X\). However both the semidifferentials are contained within (different) polyhedra defined via \(\tilde{g}_{X}\). An illustration of this is in Figure 8. The subdifferential \(\partial_{f}(X)\) is the red polyhedron, while the superdifferential \(\partial^{f}(X)\) is the blue polyhedron. Moreover, the light red and the light blue polyhedra are \(\partial_{f}^{f,(1,1)}(X)\) and \(\partial^{f,(1,1)}(X)\) respectively, defined on \(X = \{1\}\).

\(^5\)Given two polyhedra \(P_{1}, P_{2}\), \(P = \text{conv}(P_{1}, P_{2}) = \{ \lambda x_{1} + (1 - \lambda)x_{2} : \lambda \in [0, 1], x_{1} \in P_{1}, x_{2} \in P_{2} \}\)
Figure 8: An illustration to compare the relative positions of the sub and super differentials, defined on a submodular function $f : 2^{\{1,2,3\}} \rightarrow \mathbb{R}$, defined as $f(\emptyset) = 0, f(\{1\}) = 1, f(\{2\}) = 2, f(\{3\}) = 2, f(\{1,2\}) = 2.5, f(\{2,3\}) = 3, f(\{1,3\}) = 2.8, f(\{1,2,3\}) = 3$. The sub differentials appear in red, while the superdifferential is shown in blue, and are defined on $X = \{1\}$. 
Figure 9: A visualization of the inner and outer bounds of the superdifferential. The submodular function here is $f : 2^{\{1,2,3\}} \rightarrow \mathbb{R}$, defined as $f(\emptyset) = 0$, $f(\{1\}) = 1$, $f(\{2\}) = 2$, $f(\{3\}) = 2$, $f(\{1, 2\}) = 2.5$, $f(\{2, 3\}) = 3$, $f(\{1, 3\}) = 2.8$, $f(\{1, 2, 3\}) = 3$, and the superdifferential $\partial f(X)$ is at $X = \{1\}$.

The first figure (top left) is the submodular supergradient $\hat{\gamma}_f(\emptyset)$, while the second one (top) is the outer bound $\partial f(\Delta(1, 1))(\emptyset)$. The bottom three figures show the inner bounds $\partial f_{i, 1}(\emptyset)$, $\partial f_{i, 2}(\emptyset)$, and $\partial f_{i, 3}(\emptyset)$ (marked as inner bounds 1, 2 and 3 respectively).

6.2.4 Examples of Inner and Outer bounds for specific superdifferentials

In this section, we investigate the inner and outer bounds of specific instances. First consider the super differential at the emptiest $\partial f(\emptyset)$. In this case, notice that all three supergradients are the same vector, i.e $\hat{\gamma}_f(\emptyset) = \hat{\gamma}_{\emptyset} = \bar{\gamma}_f(\emptyset)$, with individual elements being $\hat{\gamma}_f(j) = \hat{\gamma}_{\emptyset}(j) = \bar{\gamma}_f(j) = f(j), j \in V$. Hence the inner bounds are exactly the superdifferential, and $\partial f_{i, 1}(\emptyset) = \partial f_{i, 2}(\emptyset) = \partial f_{i, 3}(\emptyset) = \partial f(\emptyset)$. Moreover, also observe that $\tilde{\gamma}_f$ is also identical to these supergradients, and hence the outerbound $\partial f_{\Delta(1,1)}(\emptyset) = \partial f(\emptyset)$. In this case, all the inner and outer bounds are identical to the superdifferential. This phenomenon also occurs for the superdifferential of the grounds $\partial f(V)$. For other sets, however, this does not hold, and the relationship between the bounds can be strict.

In the below, we analyze the inner and outer bounds on the superdifferentials for specific submodular functions.

Example 6.3. Consider the case when $V = \{1, 2\}$. From the above, we know that the su-
peridifferentials \( \partial^f(0) \) and \( \partial^f(\{1,2\}) \) are simple polyhedra and the inner and outer bounds are identical to the superdifferential itself. Consider \( \partial^f(\{1\}) \). Recall from Example 6.1 that the extreme points here are \( \{f(\{1\}), f(\{2\})\} \) and \( \{f(\{1\})\{2\}), f(\{2\})\{1\}\} \) respectively. Notice that \( \bar{g} = \{f(\{1\}), f(\{2\})\} \) and \( \bar{g} = \{f(\{1\})\{2\}), f(\{2\})\{1\}\} \), and hence both these supergradients are extreme points of the superdifferential in two dimensions. Note that \( \bar{g} = \{f(\{1\})\{2\}), f(\{2\})\} \). Immediately, \( \bar{g} \) lies in the interior of the superdifferential for a strictly submodular function \( f \) (this is true since \( f(0) < f(\{1\}) \)). Hence, in this case, \( \partial^f(\{1\}) \subset \partial^f(\{2\}) \). Similarly, observe that \( \bar{g}_a = \{f(\{1\}), f(\{2\})\} \) does not belong to \( \partial^f(\{1\}) \), when \( f \) is strictly submodular, since it violates the third inequality \( f(\{1\}) - f(\{2\}) \leq f(\{1\}) - f(\{2\}) \) (note this inequality does not hold since it would mean \( f(\{2\})\{1\}) \geq f(\{1\}) \) which violates strict submodularity). Hence \( \partial^f(\{1\}) \supset \partial^f(\{2\}) \). The same phenomena is true for \( \partial^f(\{3\}) \).

We can also consider the superdifferential in the three dimensional setting, when \( V = \{1,2,3\} \). In this case, we consider an actual submodular function \( f : 2^{\{1,2,3\}} \to \mathbb{R} \), defined as \( f(\emptyset) = 0, f(\{1\}) = 1, f(\{2\}) = 2, f(\{3\}) = 2, f(\{1,2\}) = 2.5, f(\{2,3\}) = 3, f(\{1,3\}) = 3, f(\{2,3\}) = 3 \). Consider \( \partial^f(\{1\}) \). An illustration of this is in Figure 6. The inner bounds \( \partial^f_{i,1}(X) \) and \( \partial^f_{i,2}(X) \) are shown in dark blue, and the superdifferential itself is shown in light blue. The outer white region is the outer bound \( \partial^f_{\Delta(1,1)}(X) \). In this case, it holds that \( \partial^f_{i,3}(\{1\}) \subset \partial^f_{i,1}(\{1\}) \subset \partial^f_{i,2}(\{1\}) \subset \partial^f(\{1\}) \subset \partial^f_{\Delta(1,1)}(\{1\}) \) and \( \partial^f_{i,3}(\{1\}) \subset \partial^f_{i,2}(\{1\}) \subset \partial^f(\{1\}) \subset \partial^f_{\Delta(1,1)}(\{1\}) \) - i.e in other words, the subset relationship is strict in this case.

### 6.2.5 Superdifferentials of subclasses of submodular functions

While it is hard to characterize superdifferentials of general submodular functions, certain subclasses have some nice characterizations. An important such subclass if the class of \( M^2 \)-concave functions [40]. These include a number of special cases like matroid rank functions, concave over cardinality functions etc. In some sense, these functions very closely resemble concave functions. In particular, one can maximize these functions in polynomial time [40]. These functions also admit simple characterizations of the superdifferential. In particular, the superdifferential of this class of functions can be represented in \( O(n^2) \) inequalities. The following theorem provides a compact representation of the superdifferential of these functions.

**Lemma 6.10.** Given a submodular function \( f \) which is \( M^2 \)-concave on \( \{0,1\}^V \), its superdifferential satisfies,

\[
\partial^f(X) = \partial^f_{\Delta(2,2)}(X)
\]

In particular, it can be characterized via \( O(n^2) \) inequalities.

**Proof.** A set function \( \mu \) is said to be \( M^2 \) [21] concave, if for any \( X, Y \subseteq V \) and \( i \in X \setminus Y \), we have,

\[
\mu(X) + \mu(Y) \leq \mu(X \setminus i) + \mu(Y \cup i)
\]

or else,

\[
\mu(X) + \mu(Y) \leq \mu(X \setminus i \cup j) + \mu(Y \cup i \setminus j)
\]

\( ^6 \)A strict submodular function, is a submodular function, where all the defining inequalities act as equalities.
for some \( j \in Y \setminus X \). This is called the exchange property.

We then invoke Theorem 6.61 in [40], where the authors show that for a \( M^2 \) convex function (which is supermodular), the subdifferential can be expressed by just considering sets \( Y \) satisfying \(|X \setminus Y| \leq 1, |Y \setminus X| \leq 1\) (i.e. of Hamming distance less than 2). In particular, we have that,

\[
\partial^\mu(X) = \{ x \in \mathbb{R}^n : x(j) \leq \mu(j|X\setminus j), \forall j \in X \}
\]

\[
x_j \geq \mu(j|X), \forall j \notin X
\]

\[
x_i - x_j \leq \mu(X) - \mu(X \cup j \setminus i), \forall i \in X, j \notin X
\]

Hence the superdifferential of a \( M^2 \) concave function (which is submodular) can be expressed with the same number of inequalities, and the corresponding polyhedron is \( \partial^\Delta_{(2, 2)}(X) \).

\[\square\]

6.3 Generalized Submodular Upper Polyhedron

In this section, we generalize the submodular upper polyhedron from Section 6.1. Define the generalized submodular upper polyhedron as the set of affine upper bounds of \( f \):

\[
P^f_{\text{gen}} = \{(x, c) , x \in \mathbb{R}^n, c \in \mathbb{R} : x(X) + c \geq f(X), \forall X \subseteq V\}
\]

Again it is easy to see that \( P^f_{\text{gen}} \cap \{(x, c) : c = 0\} = \{(x, c) : x \in P^f, c = 0\} \). In other words, the slice \( c = 0 \) of the generalized submodular upper polyhedron is the submodular upper polyhedron of \( f \).

Also note that the inequality at \( X = \emptyset \) implies that \( c \geq 0 \). This polyhedron shall prove to be useful while defining the concave extensions of \( f \). The generalized submodular upper polyhedron also has interesting connections with the superdifferentials. In particular,

**Lemma 6.11.** Given a submodular function \( f \), \( (x, c) \in P^f_{\text{gen}} \) lies on a face of the polyhedron if and only if there exists a set \( X \) such that \( x \in \partial^f(X) \) and \( c = f(X) - x(X) \).

**Proof.** The proof of this Lemma is analogous to the one for the generalized submodular lower polyhedron. In particular, observe that \( (x, c) \) lies on a face of \( P^f_{\text{gen}} \) if and only if there exists a set \( X \) such that \( x(X) + c = f(X) \) and for all \( Y \subseteq V, x(Y) + c \leq f(Y) \). It directly then implies that \( x \in \partial^f(X) \) and \( c = f(X) - x(X) \).

This then implies the following corollary:

**Corollary 6.12.** Given a submodular function \( f \), a point \( (x, c) \) is an extreme point of \( P^f_{\text{gen}} \), if and only if \( x \) is an extreme point of \( \partial^f(X) \) for some set \( X \).

This implies an interesting characterization of a linear program over the generalized submodular upper polyhedron.

**Lemma 6.13.** For submodular function \( f \), and a \( y \in \mathbb{R}^n \),

\[
\min_{(x, c) \in P^f_{\text{gen}}} \langle x, y \rangle + c = \min \{ \min_{x \in \partial^f(X)} \langle x, y \rangle + f(X) - x(X) | X \subseteq V \}.
\]

**Proof.** To prove this result, we first show that \( \min_{(x, c) \in P^f_{\text{gen}}} \langle x, y \rangle + c \leq \min \{ \min_{x \in \partial^f(X)} \langle x, y \rangle + f(X) - x(X) | X \subseteq V \} \). In order to show this, observe that for any set \( X \), and point \( x \in \partial^f(X) \), \( \langle x, f(X) - x(X) \rangle \in \partial^f(X) \). Hence the the second expression can be obtained by taking only a subset of the polyhedron \( P^f_{\text{gen}} \), and hence is a upper bound. We then show that \( \min_{(x, c) \in P^f_{\text{gen}}} \langle x, y \rangle + c \geq \min \{ \min_{x \in \partial^f(X)} \langle x, y \rangle + f(X) - x(X) | X \subseteq V \} \), by invoking Corollary 6.12. The minimum on
Then, it is immediate that the only extreme points in this case are \( V \).

**Example 6.5.**

The generalized upper submodular polyhedron in this case is,

\[
P_f^{\text{gen}} = \{(x, c) \in \mathbb{R}^3 : c \geq 0,\]
\[
x_1 + c \geq f(\{1\}), \quad (81)
\]
\[
x_2 + c \geq f(\{2\}), \quad (82)
\]
\[
x_1 + x_2 + c \geq f(\{1, 2\}) \}
\]

This is not the case for the generalized submodular upper polyhedron. Consider again the example with \( V = \{1, 2\} \).

**Example 6.5.**

The generalized upper submodular polyhedron in this case is,

\[
P_f^{\text{gen}} = \{(x, c) \in \mathbb{R}^3 : c \geq 0,\]
\[
x_1 + c \geq f(\{1\}), \quad (81)
\]
\[
x_2 + c \geq f(\{2\}), \quad (82)
\]
\[
x_1 + x_2 + c \geq f(\{1, 2\}) \}
\]

Unfortunately, however, the generalized submodular upper polyhedron is no longer easy to characterize. This is related to the fact that the superdifferentials of a submodular function are not easy to characterize.

**Lemma 6.14.** The generalized submodular upper membership problem for a submodular function \( f \) – i.e given \( x \in \mathbb{R}^n, c \in \mathbb{R} \), the problem whether \( (x, c) \in P_f^{\text{gen}} \) is NP hard for \( c > 0 \). Furthermore, for any \( y \in \mathbb{R}^n \), solving a linear program over this polyhedron, i.e \( \min_{(x, c) \in P_f^{\text{gen}}} (x, y) + c \) is also NP hard.

**Proof.** The first part of the result follows from the fact that asking whether \( (x, c) \in P_f^{\text{gen}} \) is equivalent to asking whether \( \max_{X \subseteq V} [f(x) - x(X) - c] \leq 0 \), which can be rewritten as \( \max_{X \subseteq V} [f(x) - x(X)] \leq c \). This is the decision version of submodular maximization, which is NP hard. The second part follows directly from the first since the membership problem on a polyhedron is equivalent to a linear program over this polyhedron [22, 49].

We can also prove the second part (that solving a linear program over the generalized submodular polyhedron is NP hard) since it is equivalent to computing the concave extension of a submodular function (we show this in Lemma 7.1). Computing, and in fact even evaluating at a point, this concave extension, however, is NP hard [8, 49].

Recall that in the case of the generalized submodular lower polyhedron, the extreme points of this polyhedron were identical to the extreme points of the submodular lower polyhedron (i.e all extreme points of the generalized submodular lower polyhedron occurred when \( c = 0 \)), which meant that the linear program over the two polyhedra was the same. This is not the case in the generalized submodular upper polyhedra. To see this, consider a simple example with \( V = \{1, 2\} \).

**Example 6.4.** First consider the generalized submodular lower polyhedra when \( V = \{1, 2\} \).

\[
P_f^{\text{gen}} = \{(x, c) \in \mathbb{R}^3 : c \leq 0,\]
\[
x_1 + c \leq f(\{1\}), \quad (77)
\]
\[
x_2 + c \leq f(\{2\}), \quad (78)
\]
\[
x_1 + x_2 + c \leq f(\{1, 2\}) \}
\]

Then, it is immediate that the only extreme points in this case are \( \{f(\{1\}), f(\{2\}) \}, 0 \} \) and \( \{f(\{1\}|\{2\}), f(\{2\}), 0 \} \), which are obtained by setting Eqns (77), (78), (80) and Eqns (77), (79), (80) as equalities. The extreme points in this case, are a direct product between the extreme points of \( P_f \) and \( c = 0 \). Hence all extreme points lie on the face \( c = 0 \).

This is not the case for the generalized submodular upper polyhedron.

25
Figure 10: The top figure shows the generalized submodular upper polyhedron alone (in blue), while the bottom one shows both the generalized upper and lower submodular polyhedron (the lower one is shown in red), for a submodular function $f : 2^{\{1,2\}} \rightarrow \mathbb{R}$, with $f(\emptyset) = 0, f(\{1\}) = 1, f(\{2\}) = 2, f(\{1,2\}) = 2.5$. Notice that all the extreme points of the generalized submodular lower polyhedron are on the plane $c = 0$ – the two extreme points are $\{1,1.5,0\}$ and $\{0.5,2,0\}$. In the generalized submodular upper polyhedron, however, one of the extreme points is on $c = 0$ (this extreme point is $\{1,2,0\}$), while the other extreme point is $\{0.5,1.5,0.5\}$ (here $c > 0$).
Then, it is immediate that the only extreme points in this case are \( \{f(\{1\}), f(\{2\}), 0\} \) and \( \{f(\{1\} \{2\}), f(\{2\} \{1\}), f(\{1\}) + f(\{2\}) - f(\{1, 2\})\} \), which are obtained by setting Eqns (81), (82), (83) and Eqs (82), (83), (84) as equalities (setting the other combination of inequalities as equalities does not give extreme points). Hence while one of the extreme points here is \( \{f(\{1\}), f(\{2\}), 0\} \), which is the direct product between \( P^f \) and \( c = 0 \), the other extreme point occurs at \( c > 0 \) (when \( f \) is strictly submodular).

An illustration of the generalized submodular upper and lower polyhedra is shown in Figure 10.

6.3.1 Inner and outer bounds on the generalized submodular upper polyhedron

In a manner similar to the superdifferential, we can provide inner and outer bounds of the generalized submodular upper polyhedron. In particular, let \( g^X \in \partial^f (X) \) be a supergradient, that is easy to obtain. Then, \( m^X(Y) = f(X) + g^X(Y) - g^X(X) \) is a modular upper bound of \( f(Y) \), \( \forall Y \subseteq V \). Given a set \( \{g^X | X \subseteq V\} \) of such supergradients, we may define a polytope:

\[
P^f_{g, gen} = \text{conv-hull}\{(g_Y, f(Y) - g_Y(Y)), \forall Y \subseteq V\}
\]

(85)

It is not hard to see that \( P^f_{g, gen} \subseteq P^f_{gen} \) since it is formed using some specific supergradients. Moreover, many of these bounds can be combined together by taking the convex hull of these polyhedra. We shall be interested by the polytopes formed by the supergradients formed by the supergradients \( \hat{g}_Y, \tilde{g}_Y \) and \( \check{g}_Y \) (which we call \( P^f_{1, gen}, P^f_{2, gen} \) and \( P^f_{3, gen} \)) and supergradients related to these. These bounds, as we shall see, have interesting connections to concave extensions (which we shall describe Section 7) and ultimately to submodular maximization. One can similarly also define outer bounds of the generalized submodular upper polyhedron by considering only a subset of inequalities defining \( P^f_{gen} \). We do not pursue this here, however.

7 Concave extensions of a submodular function

Following the characterizations of the convex extensions of a submodular function, we can define the concave extensions also from two viewpoints, one in the distributional setting and another in the polyhedral setting. These results follow in the lines of the results for the convex extensions.

7.1 Polyhedral characterization of the concave extension

Similar to the convex extension, the concave extension of any set function (not necessarily submodular) can be seen as the pointwise supremum of concave functions which lower bound the set function [8]. More clearly, let

\[
\Psi_f = \{\psi : \psi \text{ is concave in } [0, 1]^V \text{ and } \psi(1_X) \geq f(X), \forall X \subseteq V\}.
\]

(86)

Then define:

\[
\hat{f}(w) = \min_{\psi \in \Psi_f} \psi(w).
\]

(87)

Following arguments similar to the convex extension, eqn. (87) can be expressed as a linear program over the generalized submodular upper polyhedron.
Lemma 7.1. The concave extension in Eqn. [87] in any set function $f$ can be expressed as:

$$\hat{f}(w) = \min_{(y,c) \in \mathcal{P}_{\text{gen}}^{f}} \langle y, w \rangle + c, \forall w \in [0, 1]^{|V|}$$  \hspace{1cm} (88)

Proof. The proof of this Lemma follows along the lines of the proof of Lemma 3.1. For a given $w$, let $\psi$ be an argmin in Eqn. (87). Then since $\psi$ is a concave function in $[0, 1]^{|V|}$, there exists a supergradient $x \in \mathbb{R}^n$ at $w$ and value $d$, such that $\langle x, y \rangle + d \geq \psi(y), \forall y$ and $\langle x, w \rangle + d = \psi(w)$. In other words, $\langle x, y \rangle + d$ is a linear upper bound of $\psi(y)$, tight at $w$. Hence $f(w) = \langle x, w \rangle + d$. Finally notice that $(x, d) \in \mathcal{P}_{\text{gen}}^{f}$ since $x(X) + d \geq \psi(1_X) \geq f(X), \forall X \subseteq V$. 

Unlike the case of the convex extension, however, this is not equivalent to an optimization over the submodular upper polyhedron. Moreover, this expression requires solving a linear program over the submodular upper polyhedron, it follows from Theorem 6.14 that obtaining the concave extension is NP hard. We shall revisit this result, in the next subsection, while investigating the distributional characterization.

Interestingly, we can define a number of concave extensions based on relaxations of the polyhedral representation. In particular, consider the inner approximations of the generalized submodular upper polyhedron $\mathcal{P}_{\text{gen}}^{f}$, defined via a particular supergradient $g$. Instead of minimizing over all affine upper bounds, we can minimize only over a particular class of modular upper bounds. Then, we can define the following form of a concave extension:

$$\hat{\bar{f}}(w) = \min_{(y,c) \in \mathcal{P}_{\text{gen}}^{f}} \langle y, w \rangle + c = \min_{Y \subseteq V} \{ \langle y, g_Y \rangle + f(Y) - g_Y(Y) \}, \forall w \in [0, 1]^{|V|}$$  \hspace{1cm} (89)

In particular, the above turns the linear program into a discrete optimization problem. Moreover, the concave extension $\hat{\bar{f}}$ is guaranteed to be an upper bound of $\hat{f}$. We can define three variants of these extensions using the supergradients $\hat{g}_Y, \bar{g}_Y$ and $\hat{g}_Y$, which we call $\hat{f}_1$, $\hat{f}_2$ and $\hat{f}_3$. It is easy to see that the concave extension $\hat{f}_3$ can in fact, be obtained in polynomial time, since it involves submodular function minimization.

The class of concave extensions from Eqn. (89) has some very interesting connections to a form of concave extension proposed in [49] for monotone submodular functions. In particular, where [49] defined a concave function $\hat{f}_g$:

$$\hat{\bar{f}}(x) = \min \{ \langle f(Y) + \sum_{j \in V} x(j) f(j|Y) \rangle | Y \subseteq V \} = \min \{ \langle f(Y) + \sum_{j \not\in Y} x(j) f(j|Y) \rangle | Y \subseteq V \}$$  \hspace{1cm} (90)

This extension, can be seen as a special case of Eqn. (89) with a particular set of supergradients $g_Y$ for a monotone submodular function, defined as:

$$g_Y(j) = \begin{cases} 0 & \text{if } j \in Y \\ f(j|Y) & \text{if } j \not\in Y \end{cases}$$  \hspace{1cm} (91)

This supergradient is related to the supergradient $\hat{g}_Y$ except replacing the elements for $j \in Y$ with 0. For a monotone submodular function, this continues to remain a supergradient. This form of concave extension is NP hard to evaluate (see Section 3.7 in [49]). This shall still be useful in obtaining approximate maximizers in certain special cases (we investigate this in Section 7.4).
7.2 Distributional characterization of the concave extension

An alternate and equivalent characterization can be provided through the distributional lens.

**Lemma 7.2.** Denote \( \Lambda_w \) as the set:

\[
\Lambda_w = \{ \lambda_S, S \subseteq V : \sum_{S \subseteq V} \lambda_S 1_S = w; \sum_{S \subseteq V} \lambda_S = 1 \}
\]

(92)

The concave extension from Eqn. (88) then can also be represented as:

\[
\hat{f}(w) = \max_{\lambda \in \Lambda_w} \lambda \sum_{S \subseteq V} f(S)
\]

(93)

The proof of the above follows on similar lines as the convex extension, and is shown in [8]. Unfortunately, unlike the convex extension, this extension is NP hard to evaluate and optimize over.

**Proposition 7.3.** Given a submodular function \( f \), it is NP hard to evaluate and optimize \( \hat{f} \).

This result was shown in [49].

Similar to the polyhedral characterization, we can relax the distributional characterization to consider specific distributions. In particular, we can obtain the multilinear extension, through a particular distribution, \( \lambda_X = \prod_{i \in X} x_i \prod_{i \notin X} (1 - x_i) \).

\[
\tilde{f}(x) = \sum_{X \subseteq V} f(X) \prod_{i \in X} x_i \prod_{i \notin X} (1 - x_i)
\]

(94)

It is not hard to see that this forms a lower bound on the concave extension \( \hat{f} \). This extension, unlike most extensions seen earlier, however, is not concave. Similar to the concave extension it is hard to evaluate this extension, and typically requires sampling [49] although special cases exist where it can be evaluated analytically [31].

7.3 Concave extensions of subclasses of submodular functions

While the concave extension \( \hat{f}(x) \) is NP hard to compute in general, it can be done efficiently for certain subclasses of submodular functions. These include, for example, sums of weighted matroid rank functions [49], and the class of \( M^2 \)-concave functions (c.f. Theorem 6.42 in [40]).

7.4 Concave extensions and submodular maximization

The concave extensions and the multilinear extension have interesting connections to submodular maximization. The following lemma from [49] connects many of these extensions:

**Lemma 7.4.** [49] For every monotone submodular function \( f \), \( \hat{f}(x) \geq \tilde{f}(x) \geq (1 - \frac{1}{e}) \hat{f}(y) \).

It is also easy to relate all the three extensions of a submodular function, viz. the convex extension, the concave extension and the multilinear extension.

**Lemma 7.5.** Given a submodular function, it holds that

\[
\hat{f}(x) \geq \bar{f}(x) \geq \tilde{f}(x)
\]

(95)
Proof. The proof of this result follows directly from the distributional characterization of the convex and concave extensions. Note that the multilinear extension is a particular distribution, the concave extension is a pointwise maximum over all distributions, while the convex extension is a pointwise minimum over these distributions. 

The facts above were used in providing a relaxation based algorithm for maximizing a subclass of submodular functions efficiently [49]. This relaxation based algorithm, maximizes the concave extension $f(x)$, which though NP hard to optimize in general, can be maximized in certain special cases. The particular special case which is considered in [49], is the class of weighted matroid rank functions for which the concave extension has a simple form. Furthermore, a simple pipage rounding trick ensures no integrality gap with respect to the multilinear extension, thus providing a $1 - \frac{1}{e}$ approximation algorithm for the problem of maximizing a monotone submodular function subject to a matroid constraint. Furthermore, later, a conditional gradient style algorithm, also called the continuous greedy algorithm [50], directly optimizes the multi-linear extension, thereby providing a general $1 - 1/e$ approximation algorithm for monotone submodular maximization subject to matroid constraints. This was later extended to the non-monotone case by [6].

8 Optimality Conditions for submodular maximization

Just as the subdifferential of a submodular function provides optimality conditions for submodular minimization, the superdifferential provides the optimality conditions for submodular maximization.

8.1 Unconstrained submodular maximization

In this section, we consider the general problem of unconstrained submodular maximization:

$$\max_{X \subseteq V} f(X)$$  \hspace{1cm} (96)

Given a submodular function, we can give the KKT like conditions for submodular maximization, in the following theorem:

**Lemma 8.1.** For a submodular function $f$, a set $A$ is a maximizer of $f$, if $0 \in \partial f(A)$.

However as expected, finding the set $A$, with the property above, or even verifying if for a given set $A$, $0 \in \partial f(A)$ are both NP hard problems (from Proposition 6.5). However thanks to the submodularity, we show that the outer bounds on the superdifferential provide approximate optimality conditions for submodular maximization. Moreover, unlike the superdifferential, these bounds are easy to obtain.

**Proposition 8.2.** For a submodular function $f$, if $0 \in \partial f_{(1,1)}(A)$ then $A$ is a local maxima of $f$ (that is, $\forall B \supseteq A, f(A) \geq f(B),$ $\forall C \subseteq A, f(A) \geq f(C)$). Furthermore, if we define $S = \arg\max_{X \in (A, V \setminus A)} f(A)$, then $f(S) \geq \frac{1}{3}OPT$ where OPT is the optimal value.

The above result is interesting observation, since a very simple outer bound on the superdifferential, leads us to an approximate optimality condition for submodular maximization. The local optimality condition follows directly from the definition of $\partial f_{(1,1)}(A)$ and the approximation guarantee follows from Theorem 3.4 in [11].

We can also provide an interesting sufficient condition for the maximizers of a submodular function.
**Lemma 8.3.** If for any set $A$, $0 \in \partial^f_{i,(1,2)}(A)$, then $A$ is the global maxima of the submodular function. In particular, if a local maxima $A$ is found (which is typically easy to do), it is guaranteed to be a global maxima, if it happens that $0 \in \partial^f_{i,(1,2)}(A)$.

**Proof.** This proof follows from the fact that $\partial^f_{i,(1,2)}(A) \subseteq \partial^f(A) \subseteq \partial^f_{(1,1)}(A)$. Thus, if $0 \in \partial^f_{i,(1,2)}(A)$, it must also belong to $\partial^f(A)$, which means $A$ is the global optimizer of $f$. \qed

We can also provide similar results for constrained submodular maximization as we show in the following subsection.

### 8.2 Constrained submodular maximization

We consider here a constrained submodular maximization problem, with $C$ representing a set of sets, also called combinatorial constraints.

$$\max_{X \in C} f(X) \quad (97)$$

For example $C$ could represent a cardinality constraint $\{X \subseteq V : |X| \leq m\}$, or a spanning tree, matching, s-t path etc. Another common type of constraints are matroid constraints. Denote $I$ is the independent set of a matroid $M$. Then $C = \{X : X \in I\}$ is a matroid constraint. Similarly $C = \{X \subseteq V : c(X) \leq B\}$ represents a knapsack constraint.

Then in this case we modify the definition of $\partial^f_C(A)$ as follows:

$$\forall A \in C, \partial^f_C(A) = \{x \in \mathbb{R}^n : f(X) - f(A) \leq x(X) - x(A), \forall X \in C\} \quad (98)$$

In other words we only consider the sets belonging to the constraints. Then we can trivially define a KKT like optimality condition for this problem:

**Lemma 8.4.** For a submodular function $f$, a set $A$ is a maximizer of the problem $\max_{X \in C} f(X)$, if $0 \in \partial^f_C(A)$.

Clearly finding the set above is NP-hard. However, similar to the unconstrained setting, we show that, in a number of cases, approximating the superdifferential can lead to polynomial time algorithms for constrained submodular maximization with worst case approximation guarantees. Consider the following scenarios:

#### 8.2.1 Constrained Monotone Submodular function Maximization

Consider here a case where $f$ is a monotone submodular function, and $C$ is the constraint that the set belongs to the intersection of the independence sets of $k$ matroids. Let $M_1, M_2, \cdots, M_k$ represent $k$ matroids, with independence sets $I_1, I_2, \cdots, I_k$. Then $C = \{X : X \in \cap_{i=1}^k I_i\}$. Again, analogous to the superdifferential $\partial^f_C$, we can also define the outer-bounds $\partial^f_{C,(k,1)}$ restricted to $C$.

Then we have the following observation, for the problem of monotone submodular maximization subject to matroid constraints.

**Observation 8.1.** Given a monotone submodular function $f$ and a constraint $C = \cap_{i=1}^k I_i$, for any set $A \in C$,

1. if $0 \in \partial^f_{C,(2,k+1)}(A)$, then $f(A)$ is guaranteed to be at least $\frac{1}{k+1}$ times the optimal value. In particular, for the special case of monotone submodular maximization subject to a single matroid constraint, for any set $A \in C$, if $0 \in \partial^f_{C,(2,2)}(A)$, then $f(A)$ is guaranteed to be at least $\frac{1}{2}$ times the optimal value.
2. If $k = 1$ and $C$ is a cardinality (uniform matroid constraint $\{X : |X| \leq m\}$), for any $r > 0$, a set $A$ satisfying $0 \in \partial^f_{C,(r+1,r+1)}$ is guaranteed to have an approximation guarantee no worse than $\frac{m}{2m-r}$.

The first part of the observation (i.e point 1) follow directly from Corollary 2.4 in [36]. In the case of $k = 1$, the same result was shown in [14]. Moreover, it was also shown in [14], that for the problem of monotone submodular maximization subject to $k > 1$ matroid constraints, the approximate optimality conditions $0 \in \partial^f_{C,(2,2)}(A)$, can be arbitrarily bad, thus requiring ‘higher order’ optimality conditions. We also remark that when $r = 1$ (i.e submodular maximization subject to a single matroid constraint), this the same approximation factor can be obtained by the simple greedy algorithm [43]. The second part of the above observation (which is submodular maximization subject to cardinality constraints) follows from Theorem 8 in [13]. In the case when $r = 1$, the condition $0 \in \partial^f_{C,(2,2)}$ provides a guarantee of $m/(2m-1)$ which is a slight improvement of $1/2$ in the special case of cardinality constraints. An interesting observation is that with better forms of local optima (i.e the condition $0 \in \partial^f_{C,(r+1,r+1)}$, is a local optima upto size $r$), imply better approximation guarantees to this problem. The main insight in these results, is that the local optima which are obtained through the local search algorithms, can all be viewed as approximate optimality conditions on the superdifferential of a submodular function.

The approximation factor for $k \geq 2$ matroids can actually be improved as shown in [37].

**Observation 8.2.** Given a maximization problem of a monotone submodular function $f$ subject to $k > 1$ matroid constraints, for a set $A \in \mathcal{C}$, if $0 \in \partial^f_{C,(p+1,kp+1)}(A)$, then $f(A)$ is guaranteed to be atleast $\frac{1}{k+1/p}$ times the optimal value. In particular, for a special case of monotone submodular maximization subject to 2 matroid constraints, for any set $A \in \mathcal{C}$, if $0 \in \partial^f_{C,(p+1,2p+1)}(A)$, then $f(A)$ is guaranteed to be atleast $\frac{1}{2+1/p}$ times the optimal value.

This result is the best known result for $k > 1$ matroids and it follows from Corollary 3.1 in [37]. Similar to the monotone setting, the observation here is that the local optimality conditions from [37], can be viewed easily as approximate optimality conditions on the superdifferential of $f$.

### 8.2.2 Constrained Symmetric Submodular function Maximization

Finally we consider the case of non-monotone submodular maximization subject to $k$ matroid constraints. Consider the case of symmetric submodular functions.

**Observation 8.3.** Given a symmetric submodular function $f$,

1. If the constraint $\mathcal{C} = \bigcap_{i=1}^k I_i$, for any set $A \in \mathcal{C}$ satisfying $0 \in \partial^f_{C,(2,k+1)}(A)$, $f(A)$ is guaranteed to be atleast $\frac{1}{k+2}$ times the optimal value.

2. If $\mathcal{C}$ is the constraint that the set be a base of a matroid, any set $A$ satisfying $0 \in \partial^f_{C,(2,2)}(A)$ is guaranteed to have a valuation atleast $1/3$ of the optimal.

All the results in this proposition, follow directly from the results in [36]. The first part follows from Theorem 2.8, while the second part is implied by Theorem 5.1.

### 8.2.3 Constrained Non-monotone Submodular function Maximization

Finally, we provide an approximation bound for non-monotone submodular maximization.
We first show that a restricted version of the Discrete Separation Theorem holds. In particular, we investigate the concave variant of the Discrete Separation Theorem, and show that under another restricted form of the Discrete Separation Theorem:

\[ \text{follows by considering the functions } f \text{ and } g \text{ with } f(X) \leq g(X), \forall X, \text{ such that } f(X) \leq h(X) \leq g(X). \text{ Then the following Lemma holds:} \]

**Lemma 9.1.** Given a submodular function \( f \) and a supermodular function \( g \), such that \( f(X) \leq g(X), \forall X \subseteq V \), there exists a modular function \( h \) such that \( f(X) \leq h(X) \leq g(X) \). Moreover, when \( f \) and \( g \) are integral (and satisfy the conditions above), there exists an integral \( h \) satisfying the above.

**Proof.** Assume first that \( f(\emptyset) = g(\emptyset) \). Then Let \( h(X) = f(\emptyset) + \sum_{j \in X} f(j|\emptyset) \). Then the following chain of inequalities hold:

\[ f(X) \leq h(X) = f(\emptyset) + \sum_{j \in X} f(j|\emptyset) \leq g(\emptyset) + \sum_{j \in X} g(j|\emptyset) \leq g(X) \quad (99) \]

Since \( f(j|\emptyset) = f(j) - f(\emptyset) \leq g(j) - g(\emptyset) = g(j|\emptyset) \). The rest of the inequalities follow from submodularity (and supermodularity) of \( f \) (and \( g \)). The result for when \( f(V) = g(V) \) analogously follows by considering the functions \( f(V \setminus X) \) and \( g(V \setminus X) \) which are submodular and supermodular respectively.

The Lemma above can in fact be generalized. In particular, it is not hard to see that the result goes through whenever \( \arg\min_X [g(X) - f(X)] \) is either \( \emptyset \) or \( V \). In general, we can also provide another restricted form of the Discrete Separation Theorem:

### 9 Concave Characterizations: Discrete Separation Theorem and Fenchel Duality Theorem

In Section 5 we investigated forms of the Discrete Separation Theorem, Fenchel Duality Theorem and Minkowski Sum Theorem for submodular functions and their associated polyhedra when seen from the convex perspective. We analyze forms of the Discrete Separation Theorem and Fenchel Duality Theorem for submodular functions and their associated polyhedra, now from the concave perspective.

#### 9.1 Discrete Separation Theorem

We first show that a restricted version of the Discrete Separation Theorem holds. In particular, we investigate the concave variant of the Discrete Separation Theorem, and show that under some restricted settings, given a submodular function \( f \) and a supermodular function \( g \) with \( f(X) \leq g(X), \forall X \), there exists a modular function \( h \) such that \( f(X) \leq h(X) \leq g(X) \). Then the following Lemma holds:

**Observation 8.4.** Given a non-monotone submodular function \( f \), and \( C \) is a cardinality (uniform matroid constraint \( \{X : |X| \leq m\} \)), for any \( r > 0 \), a set \( A \) satisfying \( 0 \in \partial^f_{C,(r+1,r+1)} \) is guaranteed to have an approximation guarantee no worse than \( \frac{r}{m-r} \).

This result follows directly from Theorem 8 in [13]. The best bounds for non-monotone submodular maximization, require running several iterations of local search procedures. In particular, the procedure of [30] runs \( k + 1 \) local search procedures to obtain a \( 1/(k + 2 + 1/k) \) approximation algorithm for non-monotone submodular maximization subject to a single matroid constraint. When \( k = 1 \), running two rounds of this local search procedure results in a \( 1/4 \) approximation. The individual local search procedures here obtains a set \( A \) satisfying \( 0 \in \partial^f_{C,(2,k)}(A) \).
Lemma 9.2. Given a submodular function $f$ and a supermodular function $g$, such that $f(X) \leq g(X), \forall X \subseteq V$, let $A = \arg\min_X [g(X) - f(X)]$. Then there exists a modular function $h$ such that $f(X) \leq h(X) \leq g(X), \forall X : X \subseteq A$ or $X \supseteq A$.

Proof. The proof of this result is analogous to the earlier one. Given $A = \arg\min_X g(X) - f(X)$, assume without loss of generality that $f(A) = g(A)$. The reason, this holds without loss of generality is that, suppose that there exists a vector $h$ which separates $f$ and $g$ when $f(A) = g(A)$, the vector will continue to separate $f$ and $g$ when $f < g$. Then define $h(X) = f(A) + \sum_{j \in X \setminus A} f(j | A) - \sum_{j \in A \Delta X} f(j | A \setminus j)$. It is easy to see that $f(X) \leq h(X), \forall X \subseteq A, X \supseteq A$. Moreover, $h(X) = f(A) + \sum_{j \in X \setminus A} f(j | A) - \sum_{j \in A \Delta X} f(j | A \setminus j) \leq g(A) + \sum_{j \in X \setminus A} g(j | A) - \sum_{j \in A \Delta X} g(j | A \setminus j) \leq g(X), \forall X \subseteq A, X \supseteq A$.

The Discrete Separation Theorem may not, however, hold under the most general conditions on $f$ and $g$. However, the do hold for certain subclasses. For example, if $f$ is a $M^2$-concave function (which is submodular), and $g$ is a $M^2$-convex function (which is supermodular), the discrete separation theorem always holds (c.f. Theorem 8.15 in [10]).

9.2 Fenchel Duality Theorem

Finally we show that a version of the Fenchel duality Theorem also holds in certain restricted cases. Given a submodular function $f$ (or equivalently supermodular function $g$), define the concave Fenchel dual functions $f_\ast$ (or equivalently $g_\ast$) as:

$$f_\ast(y) = \min_{X \subseteq V} g(X) - f(X), \quad g_\ast(y) = \max_{X \subseteq V} g(X) - f(X).$$

(100)

The Fenchel duals $f_\ast$ and $g_\ast$ are concave and convex functions respectively. Unlike the convex Fenchel duals, obtaining these expressions exactly is NP hard, since it corresponds to submodular maximization. These can however be approximately obtained up to constant factors. The following Lemma then gives a restricted version of Fenchel Duality Theorem:

Lemma 9.3. Given a submodular function $f$ and a supermodular function $g$ such that the Discrete Separation Theorems hold,

$$\max_{X \subseteq V} f(X) - g(X) = \min_x g_\ast(x) - f_\ast(x)$$

(101)

Further if $f$ and $g$ are integral (and satisfy the DST), the maximum on the right hand side is attained by an integral vector $x$.

Proof. The proof of this result follows directly from Theorem 4 of [20]. In particular, [20] show that given set functions $f$ and $g$ such that the discrete separation theorem holds, the Fenchel duality theorem will also hold for this pair of functions $f$ and $g$.

Unlike the Fenchel Duality Theorem from the convex perspective, the result above may not hold in the most general setting. Moreover, if the functions $f$ and $g$ are $M^2$-concave and $M^2$-convex respectively, the Fenchel duality theorem always holds (c.f. Theorem 8.21 in [10]).

9.3 Minkowski Sum Theorem

Analogous to the results above, we show a certain restricted form of the Minkowski Sum Theorem.


Lemma 9.4. Given two submodular functions $f_1$ and $f_2$, it holds that that (the addition of the polyhedra below corresponds to a point-wise addition):

$$\mathcal{P}^{f_1 + f_2} = \mathcal{P}^{f_1} + \mathcal{P}^{f_2}$$

(102)

Similarly, $\partial^{f_1 + f_2}(\emptyset) = \partial^{f_1}(\emptyset) + \partial^{f_2}(\emptyset)$ and $\partial^{f_1 + f_2}(V) = \partial^{f_1}(V) + \partial^{f_2}(V)$.

Proof. This result follows directly from the definitions. In particular, it is easy to see that the extreme points of these polyhedra can be explicitly characterized by a submodular function. For example, the polyhedron $\mathcal{P}^{f_1}$ has a single extreme point defined by the vector $f_1(j), j \in V$. Similarly, the extreme point of $\mathcal{P}^{f_2}$ is $f_2(j), j \in V$, and the extreme point of $\mathcal{P}^{f_1 + f_2}$ is $f_1(j) + f_2(j), j \in V$, and hence the Minkowski sum theorem holds.

Unlike the Minkowski Sum theorem on the subdifferential and submodular polyhedron, the result above may not hold for the superdifferential of any arbitrary set $\partial^f(X)$ and the generalized submodular upper polyhedron $\mathcal{P}_{\text{gen}}^f$. They do hold, however, for certain subclasses of submodular functions, like the class of $M^2$-concave functions. This fact easily follows from Theorem 3 in [20], and the fact that $M^2$-concave functions satisfy the Fendel duality theorem.

10 Conclusion and Open Problems

In this manuscript, we investigated several connections between convex and concave aspects of submodular functions. We provided characterizations of the superdifferentials, concave extensions and seperation and duality theorems related to concave aspects of a submodular function, and connected these new results to existing results on the convex aspects of submodular functions. To our knowledge, this is the first work in this direction. We also show how for specific subclasses of submodular functions, like the class of $M^2$-concave set functions, this characterization is exact, while for other submodular functions, this can be done approximately.

We also leave a few open problems.

- Are there are other subclasses of submodular functions (apart from the class of $M^2$-concave set functions, for which the concave aspects, like the superdifferentials, concave extensions and characterizations like the discrete separation theorem, Fenchel duality theorem etc. can be provided exactly. In particular, we saw that the $M^2$-concave set functions, satisfy the property that $\partial^f(X) = \partial^f_{\Delta^2}(X)$. An interesting question is whether there are other interesting subclasses of submodular functions, which satisfy similar conditions on their superdifferential. Characterizing such functions, could have a direct consequence on maximizing these subclasses of submodular functions.

- In section 8 we investigated optimality conditions related to submodular maximization, and its connection to the superdifferential. An interesting open problem is if this characterization could provide insight into algorithms for submodular maximization, and conditions when submodular maximization can be done exactly. Moreover, it also is interesting that approximating the superdifferential provides different approximation algorithms for submodular maximization. It will be interesting if there is a principled relationship between these two.

- In Section 9 we study the Fenchel duality Theorem, Discrete Separation Theorem and Minkowski sum theorem, and show that these results hold under restricted settings. An open question is if Edmonds intersection theorem (cf. Section 4.1 in [19]) also holds under certain restricted settings.
• Finally, thanks to the Minkowski sum theorem, the Lovász extension of a submodular function satisfies that \( \tilde{f}_1 + \tilde{f}_2(x) = \tilde{f}_1(x) + \tilde{f}_2(x) \), i.e the Lovász extension of a sum of two submodular functions is equal to the sum of the individual Lovász extensions. An open problem is whether this relation holds (under restricted settings possibly) for the concave extensions.

References

[1] S. Ahmed and A. Atamtürk. Maximizing a class of submodular utility functions. *Math. Program., Ser. A*, 2009.

[2] F. Bach. Learning with Submodular functions: A convex Optimization Perspective (updated version). *Arxiv*, 2013.

[3] E. Boros and P. L. Hammer. Pseudo-boolean optimization. *Discrete Applied Math.*, 123(1–3):155 – 225, 2002.

[4] N. Buchbinder, M. Feldman, J. Naor, and R. Schwartz. A tight \((1/2)\) linear-time approximation to unconstrained submodular maximization. In *FOCS*, 2012.

[5] G. Calinescu, C. Chekuri, M. Pal, and J. Vondrák. Maximizing a monotone submodular function under a matroid constraint. *IPCO*, 2007.

[6] C. Chekuri, J. Vondrák, and R. Zenklusen. Submodular function maximization via the multilinear relaxation and contention resolution schemes. *STOC*, 2011.

[7] J. Djolonga and A. Krause. From MAP to Marginals: Variational Inference in Bayesian Submodular Models. In *Neural Information Processing Society (NIPS)*, Montreal, CA, December 2014.

[8] S. Dughmi. Submodular functions: Extensions, distributions, and algorithms. a survey. *arXiv preprint arXiv:0912.0322*, 2009.

[9] J. Edmonds. Submodular functions, matroids and certain polyhedra. *Combinatorial structures and their Applications*, 1970.

[10] U. Feige. A threshold of \(\ln n\) for approximating set cover. *Journal of the ACM (JACM)*, 1998.

[11] U. Feige, V. Mirrokni, and J. Vondrák. Maximizing non-monotone submodular functions. *SIAM J. COMPUT.*, 40(4):1133–1155, 2007.

[12] M. Feldman, J. Naor, and R. Schwartz. A unified continuous greedy algorithm for submodular maximization. In *FOCS*, 2011.

[13] Y. Filmus. Inequalities on submodular functions via term rewriting. *Information Processing Letters*, 2013.

[14] M. Fisher, G. Nemhauser, and L. Wolsey. An analysis of approximations for maximizing submodular set functions—ii. *Polyhedral combinatorics*, pages 73–87, 1978.

[15] A. Frank. An algorithm for submodular functions on graphs. *North-Holland Mathematics Studies*, 66:97–120, 1982.

36
[16] S. Fujishige. On the subdifferential of a submodular function. *Mathematical programming*, 29(3):348–360, 1984.

[17] S. Fujishige. Submodular systems and related topics. *Mathematical Programming at Oberwolfach II*, pages 113–131, 1984.

[18] S. Fujishige. Theory of submodular programs: A fenchel-type min-max theorem and subgradients of submodular functions. *Mathematical programming*, 29(2):142–155, 1984.

[19] S. Fujishige. *Submodular functions and optimization*, volume 58. Elsevier Science, 2005.

[20] S. Fujishige and H. Narayanan. Polyhedrally tight set functions and discrete convexity. Technical report, RIMS, Kyoto University, JAPAN, EE Dept., IIT Bombay, 2005.

[21] S. function minimization and maximization in discrete convex analysis. Murota, kazuo. Technical report, Technical Report METR 2008–32, The University of Tokyo, Japan, 2008.

[22] M. Grotschel, L. Lovász, and A. Schrijver. Geometric methods in combinatorial optimization. In *Silver Jubilee Conf. on Combinatorics*, pages 167–183, 1984.

[23] S. Iwata. A faster scaling algorithm for minimizing submodular functions. *SIAM Journal on Computing*, 32(4):833–840, 2003.

[24] S. Iwata, L. Fleischer, and S. Fujishige. A combinatorial strongly polynomial algorithm for minimizing submodular functions. *Journal of the ACM (JACM)*, 48(4):761–777, 2001.

[25] R. Iyer and J. Bilmes. Algorithms for approximate minimization of the difference between submodular functions, with applications. In *UAI*, 2012.

[26] R. Iyer and J. Bilmes. The submodular Bregman and Lovász-Bregman divergences with applications. In *NIPS*, 2012.

[27] R. Iyer and J. Bilmes. Submodular Optimization with Submodular Cover and Submodular Knapsack Constraints. In *NIPS*, 2013.

[28] R. Iyer and J. Bilmes. Submodular Point Processes. *NIPS workshop on Discrete Optimization and Machine Learning*, 2014.

[29] R. Iyer, S. Jegelka, and J. Bilmes. Curvature and Optimal Algorithms for Learning and Minimizing Submodular Functions. In *Neural Information Processing Society (NIPS)*, 2013.

[30] R. Iyer, S. Jegelka, and J. Bilmes. Fast Semidifferential based Submodular function optimization. In *ICML*, 2013.

[31] R. Iyer, S. Jegelka, and J. Bilmes. Fast Algorithms for Submodular Optimization based on Continuous Relaxations and Rounding. In *UAI*, 2014.

[32] S. Jegelka and J. Bilmes. Submodularity beyond submodular energies: coupling edges in graph cuts. In *Computer Vision and Pattern Recognition (CVPR)*, 2011.

[33] S. Jegelka and J. A. Bilmes. Notes on graph cuts with submodular edge weights. In *Neural Information Processing Society (NIPS) Workshop*, Vancouver, Canada, December 2009. Workshop on Discrete Optimization in Machine Learning: Submodularity, Sparsity & Polyhedra (DISCML).
[34] S. Jegelka, H. Lin, and J. Bilmes. On fast approximate submodular minimization. In NIPS, 2011.

[35] A. Krause and C. Guestrin. A note on the budgeted maximization of submodular functions, 2005.

[36] J. Lee, V. Mirrokni, V. Nagarajan, and M. Sviridenko. Non-monotone submodular maximization under matroid and knapsack constraints. In STOC, pages 323–332. ACM, 2009.

[37] J. Lee, M. Sviridenko, and J. Vondrák. Submodular maximization over multiple matroids via generalized exchange properties. In APPROX, 2009.

[38] H. Lin and J. Bilmes. Multi-document summarization via budgeted maximization of submodular functions. In NAACL, 2010.

[39] L. Lovász. Submodular functions and convexity. Mathematical Programming, 1983.

[40] K. Murota. Discrete Convex Analysis. Mathematical Programming, 2003.

[41] M. Narasimhan and J. Bilmes. A submodular-supermodular procedure with applications to discriminative structure learning. In UAI, 2005.

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

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

[44] R. Rockafellar. Convex analysis, volume 28. Princeton Univ Pr, 1977.

[45] S. Sahni. Computationally related problems. SIAM Journal on Computing, 3(4):262–279, 1974.

[46] A. Schrijver. Combinatorial optimization: polyhedra and efficiency, volume 24. Springer Verlag, 2003.

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

[48] Z. Svitkina and L. Fleischer. Submodular approximation: Sampling-based algorithms and lower bounds. In FOCS, pages 697–706, 2008.

[49] J. Vondrák. Submodularity in combinatorial optimization. PhD thesis, Charles University, 2007.

[50] J. Vondrák. Optimal approximation for the submodular welfare problem in the value oracle model. In STOC, pages 67–74. ACM, 2008.