Submodular Functions: Extensions, Distributions, and Algorithms
A Survey

Shaddin Dughmi

PhD Qualifying Exam Report, Department of Computer Science, Stanford University
Exam Committee: Serge Plotkin, Tim Roughgarden (Thesis Advisor), Jan Vondrák

December 16, 2009

1 Introduction

Submodularity is a fundamental phenomenon in combinatorial optimization. Submodular functions occur in a variety of combinatorial settings such as coverage problems, cut problems, welfare maximization, and many more. Therefore, a lot of work has been concerned with maximizing or minimizing a submodular function, often subject to combinatorial constraints. Many of these algorithmic results exhibit a common structure. Namely, the function is *extended* to a continuous, usually non-linear, function on a convex domain. Then, this relaxation is solved, and the fractional solution rounded to yield an integral solution. Often, the continuous *extension* has a natural interpretation in terms of distributions on subsets of the ground set. This interpretation is often crucial to the results and their analysis. The purpose of this survey is to highlight this connection between extensions, distributions, relaxations, and optimization in the context of submodular functions.

Contributions The purpose of this survey is to present a common framework for viewing many of the results on optimizing submodular functions. Therefore, most of the results mentioned – with the exception of those in Section 4.3 – are either already published, folklore, or easily gotten by existing techniques. In the first case, citations are provided. Nevertheless, for most of these results we present alternate, hopefully simplified statements and proofs that present a more unified picture. In Section 4.3, we present a new result for minimizing symmetric submodular functions subject to a cardinality constraint.

2 Preliminaries

2.1 Submodular Functions

We begin with some definitions. We consider a *ground set* X with $|X| = n$. A *set function* on X is a function $f : 2^X \to \mathbb{R}$.

Definition 2.1. *A set function* $f : 2^X \to \mathbb{R}$ *is* submodular *if, for all* $A, B \subseteq X$

$$f(A \cap B) + f(A \cup B) \leq f(A) + f(B)$$

Equivalently, a submodular function can be defined as set function exhibiting *diminishing margial returns.*

Definition 2.2. *A set function* $f : 2^X \to \mathbb{R}$ *is* submodular *if, for all* $A, B \subseteq X$ *with* $A \subseteq B$*, and for each* $j \in X$*,*

$$f(A \cup \{j\}) - f(A) \geq f(B \cup \{j\}) - f(B)$$

The fact that the first definition implies the second can be easily checked by a simple algebraic manipulation. The other direction can be shown by a simple induction on $|A \cup B| - |A \cap B|$.

We distinguish additional properties of set functions that will prove useful. We say $f : 2^X \to \mathbb{R}$ is *nonnegative* if $f(S) \geq 0$ for all $S \subseteq X$. f is *normalized* if $f(\emptyset) = 0$. f is *monotone* if $f(S) \leq f(T)$ whenever $S \subseteq T$. Moreover, f is *symmetric* if $f(S) = f(X \setminus S)$ for all $S \subseteq X$.

Algorithmic results on optimizing submodular functions can often be stated in the general *value oracle* model. This encapsulates most special cases of these functions that arise in practice. In the value oracle model, access to f is via *value queries*: the algorithm may query for the value of $f(S)$ for any S.

We conclude with some more concrete examples of submodular functions that arise in practice. We say f is a *coverage function* when elements of X are sets over some other ground set Y, and $f(S) = |\cup_{U \in S} U|$. Therefore, problems such as max-k-cover problem can be thought of as maximizing a coverage function subject to a cardinality constraint of k. Another class of submodular functions is cut functions. A set function $f : 2^V \to \mathbb{Z}$ is a *cut function* of a graph $G = (V, E)$ if $f(U)$ is the number of edges of G crossing the cut $(U, V \setminus U)$. This can be generalized to hypergraphs. Moreover, weighted versions of both coverage functions and cut functions are also submodular. There are many other examples of submodular functions, for which we refer the reader to the thorough treatment in [6].

2.2 Polytopes and Integrality Gaps

A set $P \subseteq \mathbb{R}^n$ is a *polytope* if it is the convex hull of a finite number of points, known as the *vertices* of P, in $\mathbb{R}^n$. Equivalently, $P \subseteq \mathbb{R}^n$ is a polytope if an only if it is the intersection of a finite number of halfspaces in $\mathbb{R}^n$. Polytopes are convex sets, and are central objects in combinatorial optimization.

We will consider optimizing continuous functions over polytopes. In the context of maximization problems, we say a function $F : D \to \mathbb{R}$ with $P \subseteq D \subseteq \mathbb{R}^n$ has *integrality gap* α relative to polytope P if

$$\frac{\max\{F(x) : x \in P\}}{\max\{F(x) : x \in P, x \in \mathbb{Z}^n\}} = \alpha$$

If F has integrality gap 1 relative to P, we say it has no integrality gap relative to P.

2.3 Matroids

A *Set System* is a pair (X, I), where X is the *ground set*, and I is a family of subsets of X. A special class of set systems, known as *matroids*, are of particular interest. When $M = (X, I)$ is a matroid, we refer to elements of I as the *independent sets* of M.

Definition 2.3. *A* matroid *is a set system* (X, I) *that satisfies*

- *Downwards Closure: If $T \in I$ and $S \subseteq T$ then $S \in I$.*
- *Exchange Property: If $S, T \in I$, and $|T| > |S|$, then there is $y \in T \setminus S$ such that $S \cup \{y\} \in I$.*

Given a matroid $M = (X, I)$, we define the *matroid polytope* $P(M) \subseteq [0,1]^X$ as the convex hull of the indicator vectors of the independent sets of M.

$$P(M) = hull\left(\left\{\vec{\mathbf{1}}_S : S \in I\right\}\right)$$

Edmonds [2] showed that an equivalent characterization of $P(M)$ can be given in terms of the *rank function* of the matroid. The rank function $r_M : 2^X \to \mathbb{Z}$ of matroid M is the integer-valued submodular function defined by

$$r_M(S) = \max\{|T| : T \in I, T \subseteq S\}$$

Using the rank function, the matroid polytope can be equivalently characterized as follows. For a vector $x \in \mathbb{R}^X$ and $S \subseteq X$, we use $x(S)$ to denote $\sum_{i \in S} x_i$.

$$P(M) = \{x \in \mathbb{R}+^n : x(S) \leq r_M(S) \text{ for all } S \subseteq X\}$$

We note that the vertices of the matroid polytopes are all integers, by the first definition.

3 Extensions and Distributions

An *extension* of a set function $f : 2^X \to \mathbb{R}$ is some function from the hypercube $[0,1]^X$ to $\mathbb{R}$ that agrees with f on the vertices of the hypercube. We survey various extensions of submodular functions, and connect them to distributions on subsets of the ground set.

3.1 The Convex Closure, Lovász Extension, and Chain Distributions

In this section, we will define the convex closure of any set function, and reduce minimization of the set function to minimization of its convex closure. Then, we will show that, for submodular functions, the convex closure has a simple form that can be evaluated efficiently at any point, and thus minimized efficiently.

3.1.1 The Convex Closure

For any set function $f : 2^X \to \mathbb{R}$, be it submodular or not, we can define its *convex closure* f^- as follows.

Definition 3.1. *For a set function $f : 2^X \to \mathbb{R}$, the* convex closure *$f^- : [0,1]^X \to \mathbb{R}$ is the point-wise highest convex function from $[0,1]^X$ to $\mathbb{R}$ that always lowerbounds f.*

It remains to show that convex closure exists and is well-defined. Observe that the maximum of any number (even infinite) of convex functions is again a convex function. Moreover, the maximum of any number (even infinite) of functions lowerbounding f is also a function lowerbounding f. This establishes existence and uniqueness of the convex closure, as needed. We can equivalently define the convex closure in terms of distributions on subsets of X.

Definition 3.2. *Fix a set function* $f : 2^X \to \mathbb{R}$. *For every* $x \in [0,1]^X$, *let* $D_f^-(x)$ *denote a distribution over* 2^X, *with marginals* x, *minimizing* $\mathbf{E}_{S\sim D_f^-(x)}[f(S)]$ *(breaking ties arbitrarily). The* Convex Closure f^- *can be defined as follows:* $f^-(x)$ *is the expected value of* $f(S)$ *over draws* S *from* $D_f^-(x)$.

To see that the two definitions are equivalent, let us use f_1^- and f_2^- to denote the convex closure as defined in Definitions 3.1 and 3.2 respectively. First, since the epigraph of a convex function is a convex set, it immediately follows that f_1^- lowerbounds f_2^-. Moreover, it is easy to see that $f_2^-(x)$ is the minimum of a simple linear program with x in the constraint vector; thus f_2^- is convex by elementary convex analysis. Combining these two facts, we get that $f_1^- = f_2^-$, as needed.

Next, we mention some simple facts about the convex closure of f. First, it is apparent from Definition 3.2 that the convex closure is indeed an extension. Namely, it agrees with f on all the integer points. Moreover, it also follows from Definition 3.2 that f^- only takes on values that correspond to distributions on 2^X, and thus the minimum of f^- is attained at an integer point. This gives the following useful connection between the discrete function and its extension.

Proposition 3.3. *The minimum values of* f *and* f^- *are equal. If* S *is a minimizer of* $f(S)$, *then* $\vec{\mathbf{1}}_S$ *is a minimizer of* f^-. *Moreover, if* x *is a minimizer of* f^-, *then every set in the support of* $D_f^-(x)$ *is a minimizer of* f.

3.1.2 The Lovász Extension and Chain Distributions

In this section, we will describe an extension $\mathcal{L}_f : [0,1]^X \to \mathbb{R}$, defined by Lovász in [6], of an arbitrary set function $f : 2^X \to \mathbb{R}$. In the next section we will show that, when f is submodular, $\mathcal{L}_f = f^-$. We define $\mathcal{L}_f$ as follows.

Definition 3.4. *([6]) Fix* $x \in [0,1]^X$, *and let* $X = \{v_1, v_2, \ldots, v_n\}$ *such that* $x(v_1) \geq x(v_2) \geq \ldots \geq x(v_n)$. *For* $0 \leq i \leq n$, *let* $S_i = \{v_1, \ldots, v_i\}$. *Let* $\{\lambda_i\}_{i=0}^n$ *be the unique coefficients with* $\lambda_i \geq 0$ *and* $\sum_i \lambda_i = 1$ *such that:*

$$x = \sum_{i=0}^{n} \lambda_i 1_{S_i}$$

It is easy to see that $\lambda_n = x(v_n)$, *and for* $0 \leq i < n$ *we have* $\lambda_i = x(v_i) - x(v_{i+1})$, *and* $\lambda_0 = 1 - x(v_1)$. *The value of the* Lovász extension *of* f *at* x *is defined as*

$$\mathcal{L}_f(x) = \sum_i \lambda_i f(S_i)$$

We can interpret the Lovász Extension as follows. Given a set of marginal probabilities $x \in [0,1]^X$ on elements of X, we construct a particular distribution $D^{\mathcal{L}}(x)$ on 2^X satisfying these marginals. Intuitively, this distribution puts as much probability mass on the large subsets of X, subject to obeying the marginals. Therefore, the largest possible set $X = S_n$ gets as much probability mass as possible subject to the smallest marginal $x(v_n)$. When the marginal probability $x(v_n)$ of v_n has been "saturated", we put as much mass as possible on the next largest set S_{n-1}. It is easy to see that the next element saturated is v_{n-1}, after we place $x(v_{n-1}) - x(v_n)$ probability mass on S_{n-1}. And so on and so forth. Now, it is easy to see that $\mathcal{L}_f(x)$ is simply the expected value of f on draws from the distribution $D^{\mathcal{L}}(x)$.

A note on the distributions $D^{\mathcal{L}}(*)$ defining the Lovász extension. Notice, that the definition $D^{\mathcal{L}}(x)$ is *oblivious*, in that it does not depend on the particular function f. Moreover, notice that the support of $D^{\mathcal{L}}(x)$ is a *chain*: a nested family of sets. We call such a distribution a *chain distribution* on 2^X. The following easy fact will be useful later.

Fact 3.5. *The distribution $D^{\mathcal{L}}(x)$ is the unique chain distribution on 2^X with marginals x.*

3.1.3 Equivalence of Lovász Extension and Convex Closure

We will now show that, for a submodular function f, the Lovász extension and the convex closure are one and the same. This is good news, since we can evaluate the Lovász Extension efficiently at any $x \in [0,1]^X$, and moreover we can explicitly construct a distribution with marginals x attaining the value of the Lovász Extension at x. This has implications for minimization of submodular functions, as we will show in Section 4.1.

The intuition behind this equivalence is quite simple. Recall that, from Definition 3.2, the value $f^-(x)$ is simply the minimum possible expected value of f over a distribution on 2^X with marginals x. Fixing f and x, we ask the question: what could a distribution $D_f^-(x)$ attaining this minimum look like? Submodularity of f implies that f exhibits diminishing marginal returns. Therefore, subject to the marginals x, the value of f is smallest for distributions that "pack" as many elements together as possible in expectation. By definition, that is roughly what $D^{\mathcal{L}}(x)$ is doing: it packs as many elements together subject to the smallest marginal, then packs as many unsaturated elements together until the next marginal is saturated, etc.

While the above intuition is helpful, the proof is made precise by cleaner *uncrossing* arguments. To illustrate a simple uncrossing argument, consider two sets $A, B \in 2^X$ that are *crossing*: neither $A \subseteq B$ nor $B \subseteq A$. Now, consider a simple distribution D that outputs each of A and B with probability 1/2. Now, consider *uncrossing* D to form the distribution D', which outputs each of $A \cap B$ and $A \cup B$ with probability 1/2. Observe that D and D' have the same marginals, yet by direct application of Definition 2.1 we conclude that

$$\mathop{\mathbf{E}}_{S \sim D'} f(S) = \frac{1}{2}\left(f(A \cap B) + f(A \cup B)\right) \leq \frac{1}{2}\left(f(A) + f(B)\right) \leq \mathop{\mathbf{E}}_{S \sim D} f(S)$$

Therefore, starting with any distribution, we can keep uncrossing it without changing the marginals or increasing the expected value of f. To conclude that this process terminates with a chain distribution, we need a notion of progress. We make this precise in the following Lemma and subsequent Theorem.

Lemma 3.6. *Fix a submodular function $f : 2^X \to \mathbb{R}$. Let D be an arbitrary distribution on 2^X with marginals x. If D is not a chain distribution, then there exists another distribution D' with marginals x and $\mathbf{E}_{S\sim D'} f(S) \leq \mathbf{E}_{S\sim D} f(S)$, such that $\mathbf{E}_{S\sim D'} |S|^2 > \mathbf{E}_{S\sim D} |S|^2$.*

In other words, any non-chain distribution D can be uncrossed to form a distribution D' that is no worse, and is closer to being a chain distribution. The quantity $\mathbf{E}\,|S|^2$ is simply a potential function that measures progress towards a chain distribution; other choices of potential function work equally well.

Proof of Lemma 3.6. Fix f, D and x as in the statement of the Lemma. Assume D is not a chain distribution. Therefore, there exist two sets $A, B \subseteq X$ in the support of D (i.e. $\mathbf{Pr}_D[A], \mathbf{Pr}_D[B] >$

0) that are *crossing*: neither $A \subseteq B$ nor $B \subseteq A$. Assume without loss of generality that $\mathbf{Pr}_D[B] \geq \mathbf{Pr}_D[A]$. We define a new distribution D' that simply replaces draws of A and B with draws of $A \cap B$, B, and $A \cup B$, as follows.

$$\mathop{\mathbf{Pr}}_{D'}(S) = \mathop{\mathbf{Pr}}_{D}[S] \text{ for } S \notin \{A, B, A \cap B, A \cup B\}$$

$$\mathop{\mathbf{Pr}}_{D'}(A \cap B) = \mathop{\mathbf{Pr}}_{D}[A \cap B] + \mathop{\mathbf{Pr}}_{D}[A]$$
$$\mathop{\mathbf{Pr}}_{D'}(A \cup B) = \mathop{\mathbf{Pr}}_{D}][A \cup B] + \mathop{\mathbf{Pr}}_{D}[A]$$
$$\mathop{\mathbf{Pr}}_{D'}(B) = \mathop{\mathbf{Pr}}_{D}[B] - \mathop{\mathbf{Pr}}_{D}[A]$$
$$\mathop{\mathbf{Pr}}_{D'}(A) = 0$$

Notice that distribution D' simply pairs up draws of A and B from D, and replaces each such pair with a draw of $A \cap B$ and a draw of $\mathcal{A} \cup B$. It is easy to check that this does not change the marginals x. Moreover, this allows us to conclude that the difference in the expected value of f is given by:

$$\mathop{\mathbf{E}}_{S \sim D'} f(S) - \mathop{\mathbf{E}}_{S \sim D} f(S) = [\mathop{\mathbf{Pr}}_{D}[A] f(A \cap B) + \mathop{\mathbf{Pr}}_{D}[A] f(A \cup B)] - [\mathop{\mathbf{Pr}}_{D}[A] f(A) + \mathop{\mathbf{Pr}}_{D}[A] f(B)]$$

Directly applying Definition 2.1, we conclude that this quantity is at most 0. As for the change in the potential function $\mathbf{E}[|S|^2]$, we get

$$\begin{aligned} \mathop{\mathbf{E}}_{S \sim D'} |S|^2 - \mathop{\mathbf{E}}_{S \sim D} |S|^2 &= \left(\mathop{\mathbf{Pr}}_{D}[A] \cdot |A \cup B|^2 + \mathop{\mathbf{Pr}}_{D}[A] \cdot |A \cap B|^2 \right) - \left(\mathop{\mathbf{Pr}}_{D}[A] \cdot |A|^2 + \mathop{\mathbf{Pr}}_{D}[A] \cdot |B|^2 \right) \\ &= \mathop{\mathbf{Pr}}_{D}[A] \left(|A \cup B|^2 + |A \cap B|^2 - |A|^2 - |B|^2 \right) > 0 \end{aligned}$$

Where the last inequality follows from the inclusion-exclusion equation and the strict convexity of the squaring function. ☐

Now that we know we can "uncross" any non-chain distribution without increasing the expectation of f or changing the marginals, we get the Theorem.

Theorem 3.7. *Fix a submodular function $f : 2^X \to \mathbb{R}$. For any $x \in [0, 1]^X$, we can take $D_f^-(x) = D^{\mathcal{L}}(x)$ (without loss), and therefore $f^-(x) = \mathcal{L}_f(x)$. Thus $f^- = L_f$.*

Proof. Fix f and x. Let D^* be a choice for $D_f^-(x)$ maximizing $E_{S \sim D^*}|S|^2$. The maximum is attained by standard compactness arguments. We will show that D^* is a chain distribution, which by Fact 3.5 implies that $D^* = D^{\mathcal{L}}(x)$, completing the proof.

Indeed, if D^* were not a chain distribution, then by Lemma 3.6, there exists another choice D' for $D_f^-(x)$ such that $\mathbf{E}_{S \sim D'} |S|^2 > \mathbf{E}_{S \sim D^*} |S|^2$. This contradicts the definition of D^*. ☐

The above Theorem implies the following remarkable observation: The distribution minimizing a submodular function subject to given marginals can be chosen *obliviously*, since $D^{\mathcal{L}}(x)$ does not depend on the particular submodular function f being minimized. As we will see in the next section, the same does not hold for maximization.

For completeness, we conclude with a strong converse of Theorem 3.7.

Theorem 3.8. *Fix a set function $f : 2^X \to \mathbb{R}$. If $\mathcal{L}_f$ is convex then f is submodular.*

Proof. We take a non-submodular f, and show that $\mathcal{L}_f$ is non-convex. We will show that the Lovász extension makes a suboptimal choice for minimization at some $x \in [0,1]^X$: namely, $\mathcal{L}_f(x) > f^-(x)$. By Definition 3.1, $f^-(x)$ is the point-wise greatest convex extension of f. This implies that L_f is non-convex.

We now exhibit x such that $\mathcal{L}_f(x) > f^-(x)$. By Definition 2.2, there exists a set $A \subseteq X$, and two elements $i, j \notin A$, such that

$$f(A \cup \{i,j\}) - f(A \cup \{i\}) > f(A \cup \{j\}) - f(A)$$

Define $x \in [0,1]^X$ as follows: $x(k) = 1$ for each $k \in A$, and $x(i) = x(j) = 1/2$, and $x(k) = 0$ otherwise. Now it is intuitively clear that the Lovász Extension makes the *wrong* choice for minimization: it will attempt to bundle i and j together despite *increasing* marginal returns. Indeed, By Definition 3.4, the Lovász extension at x evaluates to

$$\mathcal{L}_f(x) = \frac{1}{2} f(A \cup \{i,j\}) + \frac{1}{2} f(A)$$

Now, consider the distribution D, with marginals x, defined by $\mathbf{Pr}_D[A \cup \{i\}] = \mathbf{Pr}_D[A \cup \{j\}] = \frac{1}{2}$. Since, by Definition 3.2, $f^-(x)$ lowerbounds the expectation of any distribution with marginals x, we have that

$$f^-(x) \leq \frac{1}{2} f(A \cup \{i\}) + \frac{1}{2} f(A \cup \{j\})$$

We can now combine the three above inequalities to establish $\mathcal{L}_f(x) > f^-(x)$, completing the proof.

$$2(L_f(x) - f^-(x)) \geq f(A \cup \{i,j\}) + f(A) - f(A \cup \{i\}) - f(A \cup \{j\}) > 0$$

□

3.2 The Concave Closure

The *concave closure* of any set function can be defined analogously to the convex closure. We again state the two equivalent definitions.

Definition 3.9. *For a set function $f : 2^X \to \mathbb{R}$, the* concave closure *$f^+ : [0,1]^X \to \mathbb{R}$ is the point-wise lowest concave function from $[0,1]^X$ to $\mathbb{R}$ that always upperbounds f.*

Definition 3.10. *Fix a set function $f : 2^X \to \mathbb{R}$. For every $x \in [0,1]^X$, let $D_f^+(x)$ denote a distribution over 2^X, with marginals x, maximizing $\mathbf{E}_{S \sim D_f^+(x)}[f(S)]$ (breaking ties arbitrarily). The* Concave Closure *f^+ can be defined as follows: $f^+(x)$ is the expected value of $f(S)$ over draws S from $D_f^+(x)$.*

By a similar argument to that presented in Section 3.1.1, both definitions are well-defined and equivalent.

It is tempting to attempt to explicitly characterize the distribution $D_f^+(x)$ in the same way we characterized $D_f^-(x)$. However, no such tractable characterization is possible. In fact, it is NP-hard to even evaluate $f^+(x)$, even when f is a graph cut function.

Theorem 3.11. *([1, 7]) It is NP-hard to evaluate $f^+(x)$ for an arbitrary submodular $f : 2^X \to \mathbb{R}$ and $x \in [0,1]^X$. This is true even when f is a graph cut function.*

Proof. The proof is by reduction from the NP-hard problem Max-Cut. In the max cut problem, we are given an undirected graph $G = (V, E)$, and the goal is to find a cut $(S, V \setminus S)$ maximizing the number of edges crossing the cut. Let $f(S)$ be number of edges crossing the cut $(S, V \setminus S)$.

We reduce finding the maximum non-trivial cut (with $S \neq \emptyset, V$) to the following convex optimization problem: Maximize $f^+(x)$ subject to $1 \leq \vec{\mathbf{1}} \cdot x \leq n-1$. Indeed, it is clear that this is a relaxation of the max-cut problem. The optimum is attained at an integer point x^*, since without loss of generality the trivial sets ($\emptyset$ and V) will not be in the support of any optimum distribution. Therefore, if $f^+(x)$ can be evaluated in polynomial time for an arbitrary x, then this convex optimization problem can be solved efficiently. This completes the reduction. □

Stronger hardness results are possible. In fact, it is shown in [7] that, even when f is a monotone coverage function and k is an integer, the convex optimization problem $\max \left\{ f^+(x) : \vec{\mathbf{1}} \cdot x \leq k \right\}$ is APX-hard. More generally, it is shown in [3] that it is hard to maximize general submodular functions in the value oracle model (independently of $P \neq NP$) with an approximation factor better than $1/2$.

In light of these difficulties, there is no hope of finding exact polynomial time algorithms for maximizing submodular functions in most interesting settings, using f^+ or otherwise. Therefore, we will consider another extension of submodular functions that will prove useful in attaining constant factor approximations for maximization problems.

3.3 The Multilinear Extension and Independent Distributions

3.3.1 Defining the Multilinear Extension

Ideally, since concavity is intimately tied to maximization, we could use the concave extension of a submodular function in relaxations of maximization problems. However, unlike the convex closure, the concave closure of a submodular function cannot be evaluated efficiently. Moreover, since f^+ is the point-wise lowest concave extension of f, any concave extension will have a non-trivial integrality gap relative to most interesting polytopes, including even the hypercube. In other words, any concave extension other than f^+ will not correspond to a distribution at every point of the domain $[0,1]^X$; a property that has served us particularly well in minimization problems.

In light of these limitations of concave extensions, we relax this requirement and instead exhibit a simple extension that is *up-concave*: concave in all directions $\vec{u} \in \mathbb{R}^n$ with $u \succeq 0$ (or, equivalently $u \preceq 0$). Moreover, this extension will correspond to a natural distribution at every point, and therefore will have no integrality gap on the domain $[0,1]^X$. Surprisingly, this extension will also have no integrality gap over any matroid polytope. As we will see in Section 4.2, it turns out that, under some additional conditions, up-concave functions can be approximately maximized over a large class of polytopes.

Without further a-do, we define the *multilinear extension* F of a set function f. First, we say a function $F : [0,1]^X \to \mathbb{R}$ is *multi-linear* if it is linear in each variable x_i, when the other variables $\{x_j\}_{j \neq i}$ are held fixed. It is easy to see that multilinear functions from $\mathbb{R}^X$ to $\mathbb{R}$ form a vector space. Moreover, a simple induction on dimension shows that a multi-linear function is uniquely determined by its values on the vertices of the hypercube. This allows us to define the multilinear extension.

Definition 3.12. *Fix set function $f : 2^X \to \mathbb{R}$. The* multilinear extension $F : [0,1]^X \to \mathbb{R}$ *of f is the unique multilinear function agreeing with f on the vertices of the hypercube.*

As with the Lovász extension, the multilinear extension corresponds to a natural distribution at each point $x \in [0,1]^X$, and moreover this distribution has marginals x. This distribution becomes apparent if we express F in terms of a simple basis, with each element of the basis corresponding to a vertex of the hypercube. For a set $S \subseteq X$, we define the multilinear basis function M_S as follows

$$M_S(x) = \prod_{i \in S} x_i \cdot \prod_{i \neq S} (1 - x_i)$$

Since a multilinear function is uniquely determined by its values on the hypercube, it is easy to check that any multilinear function can be written as a linear combination of the basis functions $\{M_S\}_{S \subseteq X}$, with $f(S)$ as the coefficient of M_S.

$$F(x) = \sum_{S \subseteq X} f(S) \cdot M_S(x) = \sum_{S \subseteq X} f(S) \cdot \prod_{i \in S} x_i \cdot \prod_{i \neq S} (1 - x_i)$$

Inspecting the above expression, we notice that $F(x)$ corresponds to a simple distribution with marginals at x. Let $D^i(x)$ be the distribution on 2^X that simply picks each element $v \in X$ independently with probability $x(v)$. It is clear that $Pr_{D^i(x)}[S] = M_S(x)$. Therefore, it is clear that $F(x)$ is simply the expected value of f over draws from $D^i(x)$. This gives the following equivalent definition of the multilinear extension.

Definition 3.13. *Fix a set function $f : 2^X \to \mathbb{R}$. For each $x \in [0,1]^x$, let $D^i(x)$ be the distribution on 2^X that picks each $v \in X$ independently with probability $x(v)$. The value of the multilinear extension $F : [0,1]^X \to \mathbb{R}$ at x can be defined as the expected value of f over draws from $D^i(x)$.*

$$F(x) = \mathop{\mathbf{E}}_{S \sim D^i(x)} f(S) = \sum_{S \subseteq X} f(S) \cdot \prod_{i \in S} x_i \cdot \prod_{i \neq S} (1 - x_i) \tag{1}$$

We note that, like the Lovász extension, the multilinear extension has the property of being *oblivious*: the distribution defining F at x does not depend on the set function f. The fact that, yet again, an oblivious extension lends itself particularly well to solving optimization problems is a remarkable, and arguably fundamental phenomenon.

3.3.2 Useful Properties of the Multilinear Extension

In this section, we will develop some properties of the Multilinear extension that will be useful for problems involving maximization of submodular functions. The maximization problem we consider in Section 4.2 is that of maximizing a monotone, submodular function $f : 2^X \to \mathbb{R}$ over independent sets of a matroid $M = (X, I)$.

First, we show that the multilinear relaxation of a monotone set function is also monotone.

Proposition 3.14. *([8]) If $f : 2^X \to \mathbb{R}$ is a monotone set function, then its multilinear relaxation $F : [0,1]^X \to \mathbb{R}$ is monotone. That is, whenever $x \preceq y$, we have $F(x) \leq F(y)$.*

Proof. By Definition 3.13, it suffices to show that distribution $D^i(y)$ draws a *pointwise* larger set than $D^i(x)$. We couple draws $S_x \sim D^i(x)$ and $S_y \sim D^i(y)$ in the obvious way: for each

$v \in X$ we independently draw a random variable $R(v)$ from the uniform distribution on $[0,1]$. If $0 \le R(v) \le x(v)$, then we let $v \in S_x$, otherwise $v \notin S_x$. Similarly, $v \in S_y$ if and only if $0 \le R(v) \le y(v)$. Since we have $x(v) \le y(v)$ for every $v \in X$, it is clear that, under this coupling, $S_x \subseteq S_y$ pointwise. By monotonicity of f, this implies that $F(x) \le F(y)$. ☐

Next, we will show that the multilinear extension F of a submodular f is *up-concave*: concave when restricted to any direction $u \succeq 0$ (equivalently $u \preceq 0$). In other words, it must be that for any $x \in [0,1]^X$ and $u \succeq 0$, the expression $F(x + tu)$ is concave as a function of $t \in \mathbb{R}$ over the domain of F. This is consistent with the diminishing-marginal-returns interpretation of submodularity and the independent distribution interpretation of F: weakly increasing the marginal probability of drawing each item can only result in items getting packed together into larger and larger sets (in a point-wise sense), and hence yielding diminishing marginal increases in the expected value of f. This intuition can be made precise by carefully coupling draws from $D^i(x)$ and $D^i(y)$ for some $x \preceq y$, and considering the marginal increases from transitioning from $D^i(x)$ and $D^i(y)$ to $D^i(x + \delta u)$ and $D^i(y + \delta u)$ respectively (for some $u \succeq 0$ and arbitrarility small $\delta > 0$). However, we will instead use tools from linear algebra to get a cleaner proof.

Using elementary linear algebra, up-concavity can be re-stated as a condition on the hessian matrix $\bigtriangledown^2 F(x)$ of F at x. The matrix $\bigtriangledown^2 F(x)$ is the symmetric matrix with rows and columns indexed by X, and the (i,j)'th entry corresponding to the second partial derivative $\frac{\partial^2 F}{\partial x_i \partial x_j}(x)$. Up-concavity is then the condition that

$$u^T(\bigtriangledown^2 F(x))u \le 0 \text{ for all } x \in [0,1]^X \text{ and } u \succeq 0$$

Since F is multilinear, the diagonal entries of $\bigtriangledown^2 F(x)$ are always 0. Therefore, by considering $\left\{\vec{\mathbf{1}}_{\{i,j\}}\right\}_{i,j \in X}$ as choices for u, we conclude that F is up-concave if and only if all the second partial derivatives $\frac{\partial^2 F}{\partial x_i \partial x_j}(x)$ are non-positive. Indeed, this is consistent with submodularity and the independent-distribution interpretation of F: increasing the probability of including i only results in sets that are point-wise larger, and therefore these sets would benefit less by inclusion of j as well.

Proposition 3.15. *([8]) If $f : 2^X \to \mathbb{R}$ is submodular, then its multi-linear relaxation $F : [0,1]^X \to \mathbb{R}$ is up-concave.*

Proof. By the discussion above it suffices to show that, for each $x \in [0,1]^X$, we have that $\frac{\partial^2 F}{\partial x_i \partial x_j}(x) \le 0$. Fixing x, we take the derivative of F with respect to x_i to get:

$$\frac{\partial F}{\partial x_i} = \frac{\partial}{\partial x_i} \mathop{\mathbf{E}}_{S \sim D^i(x)} f(S) = \mathop{\mathbf{E}}_{S \sim D^i(x)} [f(S \cup \{i\}) - f(S)]$$

The equality above follows immediately from the independent distribution interpretation of F, by conditioning on all events $j \in S$ for $j \neq i$ and considering the expectation of f as a function of the marginal probability of i. Using linearity of expectation and taking the derivative again in the same way with respect to j, we get

$$\frac{\partial^2 F}{\partial x_i \partial x_j} = \mathop{\mathbf{E}}_{S \sim D^i(x)} [f(S \cup \{i,j\}) - f(S \cup i) - f(S \cup j) + f(S)]$$

Using Definition 2.1, we get that this quantity is non-positive, as needed. ☐

It is clear that, since the value of F at any point corresponds to the expectation of f at a distribution on 2^X, that F has no integrality gap relative to the hypercube $[0,1]^X$. Since we will consider constrained maximization problems, it would be useful if this held for interesting subsets of the hypercube. It turns out that, for a submodular function f, a useful property that we term *cross-convexity* yields precisely such a guarantee relative to all matroid polytopes. Cross convexity means that trading off two elements i and j gives a convex function, or increasing marginal returns.

Definition 3.16. *We say a function $F : [0,1]^X \to \mathbb{R}$ is* cross-convex *if, for any $i \neq j$, the function $F^x_{i,j}(\epsilon) := F(x + \epsilon(e_i - e_j))$ is convex as a function of $\epsilon \in \mathbb{R}$.*

Cross-convexity is consistent with submodularity and the independent distribution interpretation of the multilinear relaxation. Consider independent distribution $D^i(x)$, and the associated expectation of f. It is an easy exercise to see that the probability of "collision" of i and j – that is, the probability that both are drawn by the indepenent distribution – is a *concave* function of ϵ. Since "collision" corresponds to diminishing marginal returns, or a *decrease* in the expected value of f, this means that the expectation of f is *convex* in ϵ. We make this precise in the proposition below.

Proposition 3.17. *([1]) When $f : 2^X \to \mathbb{R}$ is submodular, its multilinear extension F is cross-convex.*

Proof. Fix x and $i \neq j$. We can write $F^x_{i,j}(\epsilon)$ as:

$$F^x_{i,j}(\epsilon) = \mathop{\mathbf{E}}_{S \sim D^i(x+\epsilon(e_i - e_j))} f(S)$$

Consider the random variable $\hat{S}$: the set of elements other than i and j that are drawn from $D^i(x + \epsilon(e_i - e_j))$. We have

$$\begin{aligned} &= \mathop{\mathbf{E}}_{\hat{S}}[(x_i + \epsilon)(x_j - \epsilon) f(\hat{S} \cup \{i,j\}) \\ &+ (x_i + \epsilon)(1 - x_j + \epsilon) f(\hat{S} \cup \{i\}) \\ &+ (1 - x_i - \epsilon)(x_j - \epsilon) f(\hat{S} \cup \{j\}) \\ &+ (1 - x_i - \epsilon)(1 - x_j + \epsilon) f(\hat{S})] \end{aligned}$$

Observe that the coefficient of ϵ^2 in the above expression is $f(\hat{S} \cup \{i\}) + f(\hat{S} \cup \{j\}) - f(\hat{S} \cup \{i,j\}) - f(\hat{S})$. By Definition 2.1, this is nonnegative, which yields convexity in ϵ as needed. ☐

Consider any $x \in [0,1]^X$ and any fractional x_i, x_j. We can trade off items i and j, in the sense defined above, until one of them is integral. Cross convexity implies that the maximum point of this tradeoff lies at the extremes. Therefore, repeating this process as long as there are fractional variables, we can arrive at an integer point $x' \in \{0,1\}^X$ such that $F(x') \geq F(x)$. When the set of feasible solutions is constrained to a proper subset of the hypercube, however, this may result in an infeasible x'. Nevertheless, for well-structured matroid polytopes, a careful rounding process maintains feasibility without decreasing the objective value. This is known as *Pipage rounding*, and will be presented in Section 4.2.

4 Algorithmic Implications

In this section, we consider minimization and maximization problems for submodular functions. The algorithms we consider will make heavy use of the extensions described in Section 3.

The algorithms we consider take as input a set X with $|X| = n$, and a rational number B. For the maximization problem we consider, additional constraints are given as input; we defer details to Section 4.2. The function $f : 2^X \to \mathbb{Q}$ to be minimized or maximized is assumed to satisfy $\frac{\max_{S \subseteq X} f(S)}{\min_{S \subseteq X} f(S)} \leq B$. The algorithms we present will operate in the value oracle model. We require that the algorithms run in time polynomial in n and $\log B$, and therefore also make a polynomial number of queries to the value oracle.

4.1 Minimizing Submodular Functions

Proposition 3.3 allows us to reduce discrete optimization to continuous optimization. Namely, we reduce minimization of f to minimization of its convex closure f^-. When f^- can be evaluated efficiently, this yields an efficient algorithm for minimizing f using the standard techniques of convex optimization.

When f is submodular, $f^- = \mathcal{L}_f$. It is clear from Section 3.1.2 that the Lovász extension can be evaluated efficiently: we can explicitly construct the distribution $D^{\mathcal{L}}(x)$, which has support of size at most $n + 1$, and then explicitly compute the expected value of f over draws from $D^{\mathcal{L}}(x)$. Therefore, we can compute the minimum of a submodular function by finding the minimum of its Lovász extension.

Theorem 4.1. *([4, 6]) There exists an algorithm for minimizing a submodular function $f : 2^X \to \mathbb{R}$ in the value query model, running in time polynomial in n and $\log B$.*

4.2 Maximizing Monotone Submodular Functions Subject to a Matroid Constraint

In this section, we consider the probelm of maximizing a nonnegative, monotone, submodular function $f : 2^X \to \mathbb{R}$ over independent sets of a matroid $M = (X, I)$. We assume f is given by a value oracle as usual, and M is given by an independence oracle: An oracle that answers queries of the form: is $S \in I$? It is well known that much can be accomplished in this independence oracle model. In particular, we can use submodular function minimization, presented in Section 4.1, to get a separation oracle for the matroid polytope $P(M)$.

First, we begin where we left off in Section 3.3. Namely, we will show that we can indeed reduce maximization of f over M to maximization of the multilinear relaxation F over the polytope $P(M)$. In particular, we show that F has no integrality gap relative to $P(M)$, and the rounding can be done in polynomial time. This is known as *Pipage Rounding.*

Lemma 4.2. *([1]) Fix a submodular function $f : 2^X \to \mathbb{R}$ and its multilinear relaxation F. Fix a matroid $M = (X, I)$. For every point $x \in P(M)$, there exists an integer point $x' \in P(M)$ such that $F(x') \geq F(x)$. Therefore, F has no integrality gap relative to $P(M)$. Moreover, starting with x, we can construct x' in polynomial time.*

Proof. Recall that the rank function r_M of matroid M is an integer valued, normalized, monotone, and submodular set function. Moreover, recall that the matroid polytope is as defined in 2.3. In the ensuing discussion, we will assume that we can efficiently check whether $x \in P(M)$, and moreover we can find tight constraint when x is on the boundary of $P(M)$. Both problems are solvable by submodular function minimization.

By multilinearity, the proposition is trivial when there is only a single fractional variable. Moreover, by multilinearity we may assume without loss of generality that every fractional variable appears in at least one tight constraint of the matroid polytope.

It follows from the submodularity of r_M that the family of "tight sets", those sets $S \subseteq X$ with $x(S) = r_M(S)$, is closed under intersection and union. Therefore, we consider a *minimal* tight set T with fractional variables x_i and x_j, and trade off x_i and x_j subject to not violating feasibility (i.e. not leaving the matroid polytope $P(M)$). Observe that, by cross-convexity, we can choose the extreme point of this tradeoff so that the value of F does not decrease. Moreover, one of two types of progress is made: either an additional variable is made integral, or a new tight set T' is created that includes exactly one of i or j. It remains to show that, repeating this process so long as there are fractional variables, the second type of progress can occur consecutively at most n times. This would complete the proof, showing that after at most n^2 steps all variables are integral.

Observe that, since T was chosen to be minimal and the tight sets are closed under intersection, trading off x_i and x_j does not "untighten" any set. Therefore, this process can only grow the family of tight sets. For simplicity, we assume that at each step we choose T to be a tight set of minimum cardinality. (This assumption can be easily removed by more careful accounting.) If no variable is made integral after trading off x_i and x_j, then an additional tight set T' is created that includes exactly one of i or j. Since tight sets are closed under intersection, and tight sets are preserved, this implies that the cardinality of smallest tight set strictly decreases. Therefore, a variable must be made integral after at most n iterations, completing the proof. □

Now, it remains to show that F can be maximized approximately over $P(M)$. In fact, something even more general is true, as shown by Vondrák in [8]: Any nonnegative, monotone, up-concave function can be approximately maximized over any solvable packing polytope contained in the hypercube. Here, by *packing polytope* we mean a polytope $P \subseteq [0, 1]^X$ that is *down monotone*: If $x, y \in [0, 1]^X$ with $x \preceq y$ and $y \in P$, then $x \in P$. A polytope P is *solvable* if we can maximize arbitrary linear functions over P in polynomial (in n) time, or equivalently if P admits a polynomial time separation oracle.

Lemma 4.3. *([8]) Fix a solvable packing polytope $P \subseteq [0, 1]^X$. Fix a nonnegative, monotone, up-concave function $F : [0, 1]^X \to \mathbb{R}+$, that can be evaluated at an arbitrary point in polynomial time. Then the problem $\max \{F(x) : x \in P\}$ can be approximated to within a factor of $1 - 1/e$ in polynomial time.*

Proof Sketch. We may assume without loss of generality that $F(\vec{0}) = 0$. We let OPT denote the maximum value of F in P, and use x^* to denote the point in P attaining this optimal. Since F is not concave in all directions, usual gradient descent techniques fail to provide any guarantees. Instead, we will show a modified gradient-descent-like technique that exploits up-concavity. We will consider a particle with starting position at $\vec{0} \in P$, and slowly move the particle in *positive directions* only: directions $u \in \mathbb{R}+^n$. This restriction is not without loss: any local descent algorithm that does not "backtrack" cannot guarantee finding the optimal solution. Nevertheless, by arguments analogous to those used for the greedy algorithm for max-k-cover, we can guarantee a $1 - 1/e$ approximation. We assume the motion of the particle is a continuous process, ignoring technical details related to discretizing this process so that it can be simulated in polynomial time.

We use $x(t)$ to denote the position of the particle at time t. We interpret the position $x(t)$ of the particle as a convex combination of vertices V_P of P, with vertex $v \in P$ having coefficient $\alpha_v(t)$

$$x(t) = \sum_{v \in V_P} \alpha_v(t) \cdot v$$

Initially, $\alpha_{\vec{0}}(0) = 1$, and $\alpha_v(0) = 0$ for each vertex $v \neq \vec{0}$. So long as $\alpha_{\vec{0}}(t) > 0$, there is room for improvement in positive directions: we can replace $\vec{0}$ in the convex combination by some other vertex $z \succeq \vec{0}$. By monotonicity, this increases the value of F.

More concretely, for a small $d_t > 0$, we let $\alpha_{\vec{0}}(t + d_t) = \alpha_{\vec{0}}(t) - d_t$, and $\alpha_z(t + d_t) = \alpha_z(t) + d_t$. We keep $\alpha_v(t + d_t) = \alpha_v(t)$ for all $v \neq \vec{0}, z$. It is clear that this process must terminate when $t = 1$, since at that point the vertex $\vec{0}$ is no longer represented in the convex combination. It remains to show how to choose z at each step so that $F(x(1)) \geq (1 - 1/e)OPT$. By simple calculus, it suffices to show that z can be chosen so that $\frac{dF(x(t))}{d_t} \geq OPT - F(x(t))$. In other words, that the rate of increase in the objective is proportional to the distance from the optimal. This is analogous to the analysis of many discrete greedy algorithms, such as that for max-k-cover.

Fixing a time t, what if we choose z so as to maximize the local gain? In other words,

$$z = \operatorname*{argmax}_{z \in P} \bigtriangledown F(x(t)) \cdot z$$

Finding such a z reduces to maximizing a linear function over the matroid polytope, which can be accomplished in polynomial time. It remains to show that there exists a $z' \in P$ with $\bigtriangledown F(x(t)) \cdot z' \geq OPT - F(x(t))$.

Consider $z' = \max(x(t), x^*) - x(t)$, where the maximization is taken co-ordinate wise. We can interpret z' as the "set-wise difference" betweeen x^* and $x(t)$. Indeed, if x^* and $x(t)$ were integral indicator vectors corresponding to subsets of X, then z' is precisely the indicator vector of their set difference. The difference between any two sets in a downwards-closed set system is again in the set system. This analogy can be made precise to show that $z' \in P$ as follows: $z' \preceq x^* \in P$.

We now show that z' gives the desired marginal increase in objective. First, it is easy to see that $x(t) + z' \succeq x^*$, and therefore by monotonicity $F(x(t) + z') \geq F(x^*) = OPT$. Moreover, since F is up-concave and $z' \succeq 0$, we get that $\bigtriangledown F(x(t)) \cdot z' \geq OPT - F(x(t))$. This completes the proof. □

When F is the multilinear relaxation of f, we can evaluate F to arbitrary precision by a polynomial number of random samples [8]. Combining Lemmas 4.2 and 4.3, we get the Theorem. Technical details that compensate for the loss of approximation due to sampling are ommitted.

Theorem 4.4. *([8]) There exists an algorithm for maximizing a nonnegative, monotone, submodular function $f : 2^X \to \mathbb{R}$ given by a value oracle, over a matroid M given by an independence oracle, that achieves an approximation ratio of $1 - 1/e$ and runs in time polynomial in n and $\log B$.*

4.3 New Result: Minimizing Nonnegative Symmetric Submodular Functions Subject to a Cardinality Constraint

In this section, we consider the problem of minimizing a nonnegative symmetric submodular function subject to a cardinality constraint. First, we make the simple observation that, by submodularity, the minimum of a symmetric submodular function is always attained at $\emptyset$ and X. Therefore, as is usual when we are working with symmetric submodular functions, we consider minimization of f

over non-empty sets. Moreover, observe that, by symmetry, an upperbound of k on the cardinality is equivalent to a lowerbound of $n - k$. Therefore, we assume without loss that are minimizing f over non-empty subsets of X of cardinality at most k.

Symmetric submodular functions often arise as cut-type functions. The cut-function of an undirected graph is the canonical example. In this context, our problem is equivalent to finding the minimum cut of the graph that is sufficiently unbalanced: i.e. with smaller side having cardinality at most k. We term this problem the minimum-unbalanced-cut problem, and point out that it has obvious implications for finding small "communities" in social networks.

A slight generalization of minimum unbalanced cut was studied in [5]. There, they consider the "sourced" version, where a designated node s is required to lie in the side of the cut of interest (the side with at most k nodes). They show that this sourced-min-unbalanced-cut problem is NP-hard by reductions from at-most-k-densest subgraph and max-clique. Moreover, they give an algorithm achieving a bicriteria result parametrized by $\alpha > 1$: They find a cut of capacity at most α of the optimal unbalanced cut, yet violating the cardinality constraint by a factor of up to $\frac{\alpha}{\alpha-1}$. When $\alpha = 2$, this gives a 2-approximation algorithm that overflows the constraint by a factor of at most 2. Their techniques do not directly yield a constant approximation algorithm for the problem without violating the constraint.

4.3.1 A 2-approximation algorithm

In this section, we show a 2-approximation algorithm for minimizing a nonnegative, symmetric submodular function subject to a cardinality constraint. Without loss, we assume the constraint is an upper bound of k on the cardinality of the set. The algorithm operates in the value query model, and runs in polynomial time. This result is stronger than the result in [5] in two ways: It applies to general nonnegative symmetric submodular functions rather than just graph cut functions, and it achieves a constant factor approximation without violating the constraint. The reader may notice, however, that this problem as-stated is not strictly more general than the "sourced" problem considered in [5]. We leave open the question of whether a similar guarantee is possible for the sourced problem.

We will now argue that Algorithm 1 runs in polynomial time. Step 2 can completed in polynomial time by standard convex optimization techniques. For step 3, the polynomial-time construction in Section 3.1.2 computes an explicit representation of $D^{\mathcal{L}}(x)$. Moreover, from Section 3.1.2 we know that $D^{\mathcal{L}}(x)$ has a support of size at most $n + 1$, and thus steps 4 and 7 can be completed in polynomial time. It is then easy to see that the entire algorithm terminates in polynomial time.

Next, we argue correctness by nondeterministically stepping through the algorithm. Let S^* denote the optimal solution to the problem, with $f(S^*) = OPT$. First, assume the algorithm guesses some $v_1 \in S^*$. Since $\mathcal{L}_f$ is an extension of f and S^* has cardinality at most k, step 2 computes x with $\mathcal{L}_f(x) \le OPT$. Moreover, we know from Section 3.1.2 that $\mathcal{L}_f(x)$ is the expected value of f over draws from $D^{\mathcal{L}}(x)$.

If S with $|S| \le k$ and $f(S) \le 2L_f(x)$ is found in step 4, then we terminate correctly with a 2-approximation. Otherwise, we can show that step 7 finds S' with $f(S') \le OPT$.

Lemma 4.5. *Either there exists S in the support of $D^{\mathcal{L}}(x)$ with $|S| \le k$ and $f(S) \le 2L_f(x)$, or there exists S' in the support of $D^{\mathcal{L}}(x)$ with $|S'| \le 2k$, and $f(S') \le L_f(x)$.*

Algorithm 1 2-approximation for minimizing nonnegative, symmetric, submodular f subject to cardinality constraint.

Require: $f : 2^X \to \mathbb{R}$ a nonnegative, symmetric, submodular function given by a value oracle. Integer k such that $0 < k < n$.

Ensure: Q minimizes f over non-empty sets of size at most k

for all $v_1 \in X$ **do**
  Find $x \in [0,1]^X$ minimizing Lovász extension $\mathcal{L}_f$ subject to $x(v_1) = 1$ and $\vec{\mathbf{1}} \cdot x \leq k$.
  Construct the Lovász extension distribution $D^{\mathcal{L}}(x)$ corresponding to point x.
  if There is S in the support of $D^{\mathcal{L}}(x)$ with $|S| \leq k$ and $f(S) \leq 2L_f(x)$ **then**
    return S
  else
    Find S' in the support of $D^L(x)$ minimizing $f(S')$ subject to $|S'| \leq 2k$
    for all $v_2 \in S' \setminus \{v_1\}$ **do**
      Using submodular minimization, find T minimizing $f(T)$ subject to $v_1 \in T$ and $v_2 \notin T$.
      if $|T \cap S'| \leq k$ **then**
        $Q_{v_1,v_2} := T \cap S'$
      else
        $Q_{v_1,v_2} := \overline{T} \cap S'$
      end if
    end for
  end if
end for
$Q := \operatorname{argmin}_{v_1,v_2} f(Q_{v_1,v_2})$
return Q

Proof. Assume not. It is now easy to check that each set R in the support of $D^{\mathcal{L}}(x)$ has

$$f(R) > \left(2 - \frac{|R|}{k}\right) \mathcal{L}_f(x)$$

Taking expectations, we get that

$$\mathop{\mathbf{E}}_{R \sim D^{\mathcal{L}}(x)} f(R) > \left(2 - \mathop{\mathbf{E}}_{R \sim D^{\mathcal{L}}(x)} \frac{|R|}{k}\right) \mathcal{L}_f(x) \geq L_f(x)$$

The last inequality follows from the fact that the expected value of $|R|$ is at most k, by definitoin of $D^{\mathcal{L}}(x)$. This is a contradiction, since by definition the expectation of f over draws from $D^{\mathcal{L}}(x)$ is precisely $\mathcal{L}_f(x)$. □

Now, assuming no appropriate S was found in step 4, we have S' as in the statement of Lemma 4.5 with $k < |S'| \leq 2k$ and $v_1 \in S'$. Since $|S^*| \leq k$, we know that there exists $v_2 \in S'$ such that $v_2 \notin S^*$. In particular, there exists a set containing v_1 and not containing v_2 with value at most OPT. Assume the algorithm guesses such a v_2. This immediately yields the following Lemma.

Lemma 4.6. *If $v_1 \in S^*$ and $v_2 \in S' \setminus S^*$ then step 9 finds T such that $f(T) \leq OPT$.*

Therefore, combining Lemmas 4.5 and 4.6, we get the following from submodularity and non-negativity:

$$f(T \cap S') \leq f(T \cap S') + f(T \cup S') \leq f(T) + f(S') \leq OPT + OPT = 2OPT$$

Moreover, we know by symmetry of f that $f(\overline{T}) = f(T) \leq OPT$. Therefore, by the same calculation we get $f(\overline{T} \cap S') \leq 2OPT$. Now, observe that $T \cap S'$ and $\overline{T} \cap S'$ partition S' into non-trivial subsets by definition of T. This gives that the smaller of the two, Q_{v_1,v_2}, has cardinality between 1 and k, and moreover $f(Q_{v_1,v_2}) \leq 2OPT$. The algorithm tries all v_1 and v_2, so this immediately yields the Theorem.

Theorem 4.7. *Algorithm 1 is a polynomial-time 2-approximation algorithm for minimizing a non-negative, symmetric, submodular function subject to a cardinality constraint in the value oracle model.*

Conclusion In this survey, we considered various continuous extensions of submodular functions. We observed that those extensions yielding algorithmic utility are often associated with natural, even *oblivious* distributions on the ground set. We presented a unified treatment of two existing algorithmic results, one on minimization and one on maximization, using this distributional lens. Moreover, we demonstrate the power of this paradigm by obtaining a new result for constrained minimization of submodular functions.

Acknowledgements The author would like to thank the qualifying exam committee: Serge Plotkin, Tim Roughgarden, and Jan Vondrák, for helpful discussions. We single out Jan Vondrák for his careful guidance, and for pointing out an improvement in the approximation ratio of the result in Section 4.3.

References

[1] Gruia Calinescu, Chandra Chekuri, Martin Pál, and Jan Vondrák. Maximizing a submodular set function subject to a matroid constraint (extended abstract). In *IPCO*, pages 182–196, 2007.

[2] Jack Edmonds. Matroids, submodular functions, and certain polyhedra. In *Combinatorial Structures and Their Applications*, 1970.

[3] Uriel Feige, Vahab S. Mirrokni, and Jan Vondrák. Maximizing non-monotone submodular functions. In *FOCS*, pages 461–471, 2007.

[4] Martin Grötschel, Lázló Lovász, and Alexander Schrijver. The ellipsoid method and its consequences in combinatorial optimization. In *Combinatorica 1*, 1981.

[5] Ara Hayrapetyan, David Kempe, Martin Pál, and Zoya Svitkina. Unbalanced graph cuts. In *ESA*, pages 191–202, 2005.

[6] Lázló Lovász. Submodular functions and convexity. In *Mathematical Programming– The State of the Art (Bonn, 1982)*, 1982.

[7] Jan Vondrák. Submodularity in combinatorial optimization. In *PhD Thesis, Charles University*, 2007.

[8] Jan Vondrák. Optimal approximation for the submodular welfare problem in the value oracle model. In *STOC*, pages 67–74, 2008.