princeton univ. F'23 cos 521: Advanced Algorithm Design

Lecture 7: Submodular Functions, Lovász Extension and Minimization

Lecturer: *Huacheng Yu*

Based on the notes from Ankur Moitra's class at MIT (see references for a link).

This lecture introduces submodular functions as a generalization of some functions we have previously seen for e.g. the cut function in graphs. We will see how we can use the ellipsoid method developed in the previous lecture to minimize an arbitrary submodular function.

1 Submodular Functions

Definition 1 (Submodular Functions)**.** *Let N be a set of n elements. A function $f : 2^N \to \mathbb{R}$ is said to be submodular, if it satisfies following property of* diminishing marginal returns*: for every $A \subseteq B \subseteq N$ and $j \notin B$, $f(A \cup \{j\}) - f(A) \geq f(B \cup \{j\}) - f(B)$.*

One way to understand submodularity is to think of f as a *utility* functions. Then, the diminishing marginal returns property says that the marginal utility gained by adding a new item to a smaller set is no less than the marginal utility gained by adding a new item to a larger set.

The defining property of diminishing marginal returns is equivalent to the following (see the notes by Jeff Bilmes, in references, for a proof):

Lemma 1. *A function $f : 2^N \to \mathbb{R}$ is submodular if and only if for every $X, Y \subseteq N$, $f(X) + f(Y) \geq f(X \cap Y) + f(X \cup Y)$.*

2 Examples of Submodular Functions

Submodular functions generalize familiar quantities studied before in this class. For a graph $G(V, E)$ on n vertices, let $f(S) = E(S, \bar{S})$, the number of edges that lie in the cut defined by the vertex set S. It is a short exercise to prove that f is a submodular function.

Another important example is a *coverage function*: let $G(U, V, E)$ be a bipartite graph with the bipartition U and V of vertices with $|U| = n$. For every $S \subseteq U$, define $f(S)$ to be the size of the neighborhood of vertices in S. Then f is a submodular function. f is also *monotone* — for any S and $j \in U$, $f(S \cup \{j\}) \geq f(S)$. Observe that the cut function is not necessarily monotone.

Our final example is from information theory — let $x_1, x_2, \ldots, x_n$ be discrete random variables. For any $A \subseteq [n]$, let $H(A)$ be the joint entropy of the $\{x_i\}_{i \in A}$, i.e., $H(A) = -\sum_a P(\{x_i\}_{i \in A} = a) \log P(\{x_i\}_{i \in A} = a)$. Then, H is a submodular function.

3 Representation and Optimization

For the special cases above, we are familiar with the algorithmic tasks of optimization. For example, we know algorithms for finding the minimum size cut in a graph that run in polynomial time, along with a poly-time 0.878 approximation algorithm for the problem of finding a maximum size cut in a graph. Incidentally, while the above approximation factor requires the use of semidefinite programming, there's a simple algorithm that returns a random cut which yields a 0.5 approximation ratio. Incidentally, this trivial algorithm cannot be improved by even a sub-exponential size linear programs (see references) even though a simple semidefinite program does phenomenally better!

What about optimizing arbitrary submodular functions? This immediately leads to the question of how the submodular function we are interested in is represented. This is a tricky issue and in general, there exist submodular functions which do not have (even in an approximate sense) a small representation (see the paper by Balcan and Harvey in the references for details of one such construction). In most applications, we do have a concise representation. However, we would like our algorithms to not be tied to a particular representation, and we'd like to make minimal assumptions about it. We will see that we can give non-trivial algorithms by just having a oracle access to the submodular function — we will only need that given any x, some oracle returns to us the value of the underlying function at x.

As the case of min-cut might suggest, there's an efficient algorithm for minimizing an arbitrary submodular function given just oracle access to it! The maximization problem is of course NP-hard (Max-Cut is a special case) but it turns out that for a non-negative submodular function the oracle access is enough to yield a 0.5 approximation guarantee (this is by no means trivial though — see the paper by Buchbinder, Naor and Schwartz in references.) The naive algorithm that selects each element independently with probability half is known to give 0.25 approximation.

Our interest here is going to be in minimizing an arbitrary submodular function.

4 Minimizing Convex Functions over Convex Sets

Our main interest today is to give a polynomial time algorithm for minimizing an arbitrary submodular function. Even though submodular function is a discrete object, we will be able to minimize it by reduction to minimizing a *continuous convex function* over a convex set. Before describing this reduction, we recall how to use the ellipsoid method in order to minimize an arbitrary convex function $f : \mathbb{R}^n \to \mathbb{R}$ over a convex set K.

We will need the following assumptions:

1) There's an efficient *evaluation oracle* and *gradient oracle* for f i.e., given a point x, we can compute $f(x)$ and $\nabla f(x)$.

2) There's a known ellipsoid containing K and there's an efficient separation oracle for K.

First note that by doing a binary search over a parameter c, we can reduce minimizing f to the problem of finding a point (or proving emptiness) in $S_c = K \cap \{x \mid f(x) \leq c\}$ given a (rough) interval where the minimum must lie. Notice that S_c is convex whenever f is a convex function.

We now want to use the ellipsoid method to find a point in S_c or to prove it is empty. Recall that we can use the ellipsoid method from previous classes to do this provided we can satisfy two conditions:

a) There's an ellipsoid that contains S_c (one usually knows a bound on the diameter of the set K in which case we can just set this ellipsoid to be the sphere of the same diameter.) This is satisfied immediately from the assumption on K above.

b) There's an efficient separation oracle for S_c: We want to show that given a separation oracle for K, we can build an efficient separation oracle for S_c. Let $x \in \mathbb{R}^n$. Then, we first test if $x \in K$, and if $x \notin K$ then obtain a separating hyperplane using the separation oracle for K. In case $x \in K$, we test if $f(x) \leq c$ by using the evaluation oracle for f. If $f(x) > c$, then, observe that by convexity of f, for any minimizer of f, say x^*, $f(x) - f(x^*) \leq \nabla f(x)(x - x^*)$ which is a hyperplane (in x^* after we put c instead of $f(x^*)$) that separates x and the convex set of all minimizers of f.

Thus, we have that we can minimize f over K efficiently.

5 Lovász Extension

We want to use the above idea for minimizing a convex function given oracle + gradient access over a convex set to minimize an arbitrary submodular function. To do this, we will define an *extension* $\hat{f}$ of a given submodular function f; here, $\hat{f}$ will be a continuous function that *extends* f to all of $[0,1]^n$. The definition itself makes sense for an arbitrary set function $f : 2^N \to \mathbb{R}$. However, $\hat{f}$ will be convex if and only if f is submodular.

First, we think of $f : 2^N \to \mathbb{R}$ as $f : \{0,1\}^n \to \mathbb{R}$ by associating a set S with its 0-1 indicator from $\{0,1\}^n$ (recall that $n = |N|$).

Definition 2. *Let $f : \{0,1\}^n \to \mathbb{R}$. Define $\hat{f}$, the Lovász extension of f, so that for any $x \in [0,1]^n$, $\hat{f}(x) = \mathbb{E}_{\lambda \sim [0,1]}[f(\{i \mid x_i \geq \lambda\})]$ where $\lambda \sim [0,1]$ denotes a uniformly random sample from the interval $[0,1]$.*

Why is $\hat{f}$ an extension of f? Let $x \in \{0,1\}^n$. Then notice that for any $\lambda \in [0,1]$, $\{i \mid x_i \geq \lambda\} = S$ where $S = \{i \mid x_i = 1\}$. Thus, $\hat{f}$ agrees with f over the hypercube and is some kind of an average of f at fractional points.

We can in fact explicitly find out this averaging representation of $\hat{f}(x)$. To do this, we define a "chain" of sets associated with any $x \in \{0,1\}^n$. For simplicity of notation, assume that $x_0 = 1 \geq x_1 \geq x_2 \ldots \geq x_n \geq 0 = x_{n+1}$. Let S_i for any $i \in [n] \cup \{0\}$ equal $\{1, 2, \ldots, i\}$. Then, $S_0 = \emptyset \subseteq S_1 \subseteq S_2 \ldots \subseteq S_n = [n]$.

Further notice that if $\lambda \in [x_i, x_{i+1})$ (the probability of which is equal to $x_i - x_{i+1}$) for any $i \in [n] \cup \{0\}$, then $\{j \mid x_j \geq \lambda\} = S_i$. Thus,

$$\hat{f}(x) = \sum_{i=0}^{n} (x_i - x_{i+1}) f(S_i). \tag{1}$$

In particular, to evaluate $\hat{f}$ at any x, we need only $n+1$ calls to the evaluation oracle for f.

6 Lovász Extension is Convex iff f is submodular

The key result of today is the following:

Theorem 2. *Let $\hat{f}$ be the Lovász Extension of $f : \{0,1\}^n \to \mathbb{R}$. Then, $\hat{f}$ is convex iff f is submodular.*

We will only show the "if"part — that is, if f is submodular then $\hat{f}$ is convex. This is what is needed for our minimization algorithm.

Let us first consider a generic question about extending a function on a (discrete) set of points to a convex function. Suppose we have a set P such that every point in P cannot be represented as a convex combination of the others, and a function $f : P \to \mathbb{R}$. How do we extend f to a convex function g defined on the convex hull of P?

By the definition of a convex function, if a point x can be represented as a convex combination of points in P, then its value must be *at most* the same convex combination of the values of those points. This gives an upper bound on $g(x)$, for each way of writing x as a convex combination of P. In particular, $g(x)$ must be at most the smallest upper bound obtained in this way. It turns out that, if for every x, we set $g(x)$ to be equal to this smallest upper bound, the function we get is convex. This is called the convex closure of f.

Definition 3. *For $f : P \to \mathbb{R}$, let the convex closure of f be the function g such that*

$$g(x) = \min\left\{\sum_{z\in P} \alpha_z f(z) : \sum_z \alpha_z = 1 \wedge \sum_z \alpha_z \cdot z = x \wedge \forall z \in P, \alpha_z \geq 0\right\},$$

for all x in the convex hull of P.

Lemma 3. *g is convex.*

Proof. We prove by verifying for any x_1, x_2 in the convex hull, and $\lambda \in [0,1]$,

$$\lambda g(x_1) + (1-\lambda) g(x_2) \geq g(\lambda x_1 + (1-\lambda)x_2).$$

Suppose $g(x_1) = \sum_{z\in P} \alpha_z f(z)$ and $g(x_2) = \sum_{z\in P} \alpha'_z f(z)$. Then LHS is

$$\sum_{z\in P} (\lambda \alpha_z + (1-\lambda)\alpha'_z) f(z).$$

Note that the coefficients $(\lambda\alpha_z + (1-\lambda)\alpha'_z)$ are nonnegative, sum up to one, and satisfy

$$\sum_z (\lambda\alpha_z + (1-\lambda)\alpha'_z) z = \lambda x_1 + (1-\lambda) x_2.$$

Thus, LHS is at least $g(\lambda x_1 + (1-\lambda)x_2)$, as $g(\lambda x_1 + (1-\lambda)x_2)$ is the minimum over all such convex combinations that are equal to $\lambda x_1 + (1-\lambda)x_2$. □

Next, we show that when $P = \{0,1\}^n$ and f is submodular, $\hat{f}$ is equal to g. Hence, the Lovász extension is convex. For simplicity of notations, we may use S to denote a subset of $[n]$, and simultaneously its indicator vector in $\{0,1\}^n$. We do this in two steps:

first, we show that in the definition of g, we can always assume that for each $x \in [0,1]^n$, the minimum is achieved on some set of coefficients $\{\alpha_S\}_{S\in\{0,1\}^n}$ that takes non-zero values only on a chain of at most $n+1$ sets $S_0 \subset S_1 \cdots \subset S_n$; then we show that for every $x \in [0,1]^n$, there is essentially only a unique way to write x as a convex combination of this form, and observe that $\hat{f}$ also has this form.

Lemma 4. *When f is submodular, for every $x \in [0,1]^n$, there exists $S_0 \subset \cdots \subset S_n$ and nonnegative $\alpha_{S_0}, \ldots, \alpha_{S_n}$ summing up to 1 such that*

- $\sum_{i=0}^n \alpha_{S_i} \cdot S_i = x$;
- $g(x) = \sum_{i=0}^n \alpha_{S_i} \cdot f(S_i)$.

Proof. Fix $x \in [0,1]^n$, and let $\{\alpha_S\}_S$ be the minimizer in the definition of $g(x)$, in case of tie, choose the one that maximizes $\sum_S \alpha_S |S|^2$.[1] We show that this collection of coefficients $\{\alpha_S\}_S$ has the claimed property.

Suppose there exists $A, B \subseteq [n]$ such that $\alpha_A, \alpha_B > 0$ and $A \subsetneq B, B \subsetneq A$. Let $\Delta = \min\{\alpha_A, \alpha_B\} > 0$, and consider a set of coefficients $\{\alpha'_S\}$ such that

- $\alpha'_S = \alpha_S - \Delta$ for $S = A, B$;
- $\alpha'_S = \alpha_S + \Delta$ for $S = A \cap B, A \cup B$;
- $\alpha'_S = \alpha_S$ for $S \neq A, B, A \cap B, A \cup B$.

$\{\alpha'_S\}$ gives another convex combination of x, and by submodularity of f, $f(A) + f(B) \geq f(A \cup B) + f(A \cap B)$, thus, $\sum \alpha_S f(S) \geq \sum \alpha'_S f(S)$. In particular, $\{\alpha'_S\}$ is also a minmizer in the definition of $g(x)$. Moreover, we have $\sum_S \alpha_S |S|^2 < \sum_S \alpha'_S |S|^2$, since

$$|A \cap B|^2 + |A \cup B|^2 > |A|^2 + |B|^2 .$$

This is a contradiction.

Thus, all sets S that have nonzero α_S must form a chain $S_0 \subset \cdots \subset S_n$. $\square$

Now fix $x \in [0,1]^n$, for simplicity assume that $1 \geq x_1 \geq x_2 \cdots \geq x_n \geq 0$. Otherwise, we rename the coordinates. Let us first consider the case where $1 > x_1 > \cdots > x_n > 0$. For such x, the only way to write x as a convex combination of a chain of sets is to let $S_i = \{1, \ldots, i\}$, and $\alpha_{S_i} = x_i - x_{i+1}$. This is because for any $j \in S_{i+1} \setminus S_i$, we have $x_j = \alpha_{S_{i+1}} + \cdots + \alpha_{S_n}$, which is smaller for larger i. That means 1) since all coordinates are different, no difference $S_{i+1} \setminus S_i$ can have more than one element; 2) the element in the difference $S_{i+1} \setminus S_i$ must be increasing as i gets larger (since $x_j > x_{j+1}$). Thus, S_i has to be $\{1, \ldots, i\}$, and its coefficient α_{S_i} can be determined by taking the difference. Therefore, $g(x) = \sum_i (x_i - x_{i+1}) f(S_i) = \hat{f}(x)$.

The case for general x with possibly equal coordinates can be proved similarly, or by simply using the fact that both g and $\hat{f}$ are continuous, and are equal on all other x.

This proves that $\hat{f}$ is convex.

[1] Such $\{\alpha_S\}$ exists, since we are minimizing / maximizing a continuous function over a compact set.

7 Wrapping Up

Now that we know that the Lovász extension of f is convex iff f is submodular, we can minimize the Lovász extension of f over $[0,1]^n$ whenever f is submodular. At first glance, it's not clear what this buys us: what happens if the minimizer isn't in $\{0,1\}^n$? How do we then turn this into a minimizer for f?

The other magic property of the Lovász extension (which is especially surprising when it is convex), is that it always achieves its minimum at an extreme point (and moreover, a minimum on the extreme point can be found given any interior minimum).

To see this, consider any minimizer $\vec{x}$. For simplicity of notation, again assume that $1 \geq x_1 \geq \ldots \geq x_n \geq 0$. Then $\hat{f}(x) = \sum_{i=0}^{n}(x_i - x_{i+1})f(S_i)$ as stated previously. Therefore, $\hat{f}(x)$ is a convex combination of f evaluated at $n+1$ different sets, and $\hat{f}(x) \geq \min_i f(S_i)$. In fact, as $\vec{x}$ was the minimum, we must have $\hat{f}(x) = f(S_i)$ for all i such that $x_i - x_{i+1} > 0$. Therefore, once we know the minimizer $\vec{x}$ of $\hat{f}(\cdot)$, we can immediately output the minimizer S of $f(\cdot)$.

The "basic part" is the work immediately above: the minimizer of the Lovász extension is always an extreme point, and an extreme point minimizer can be deduced from a minimizer of the extension. Because this point is in fact a minimum even of the extension to $[0,1]^n$, it is certainly the minimum over $\{0,1\}^n$. Normally, functions which have minima on the extremes are not convex.

The "magic part" is that the Lovász extension happens to be convex when f is submodular. When f is submodular, we can actually minimize the extension in poly-time, enabling us to minimize f.

Bibliography

1. *Lecture Slides: "Submodular functions, their optimization and applications."* Jeff Bilmes. Dept. of Elect. Eng., Univ. Washington, Seattle, Apr. 1, 2011 [Online]. Available: http://melodi.ee.washington.edu/~bilmes/ee595a_spring_2011/lecture2.pdf

2. *A tight linear time (1/2)-approximation for unconstrained submodular maximization.* Niv Buchbinder, Moran Feldman, Joseph Naor, and Roy Schwartz. In FOCS, 2012.

3. *Approximating Rectangles by Juntas and a weakly exponential lower bound on linear programs for constraint satisfaction* Pravesh K. Kothari, Raghu Meka and Prasad Raghavendra. Available: https://arxiv.org/abs/1610.02704

4. *Learning Submodular Functions* Maria Florina Balcan and Nick Harvey. In STOC 2011.

5. *Lecture Notes: Submodular Functions and Lovász Extension* Ankur Moitra. Available at: http://people.csail.mit.edu/moitra/docs/6854notes13.pdf