**CS 369P: Polyhedral techniques in combinatorial optimization**
**Instructor: Jan Vondrák**
**Lecture date: November 18, 2010**

# 1 Continuous extensions of submodular functions

Submodular functions are functions assigning values to all subsets of a finite set $N$. Equivalently, we can regard them as functions on the boolean hypercube, $f : \{0,1\}^N \to \mathbb{R}$. It has been often said that submodular functions are analogous to convex or perhaps concave functions. But this analogy is somewhat nebulous and indeed it is not very clear whether submodular functions are "convex" or rather "concave". A question related to this is, what is a "natural" extension of a function $f : \{0,1\}^N \to \mathbb{R}$ to the domain $[0,1]^N$? This question does not have a unique answer.

## 1.1 Convex and concave closures

First, every function $f : \{0,1\}^N \to \mathbb{R}$ has two canonical extensions $f^+, f^- : [0,1]^N \to \mathbb{R}$, where $f^+$ is concave and $f^-$ is convex. These functions are the *concave closure* and *convex closure* of $f$.

**Definition 1** *For $f : \{0,1\}^N \to \mathbb{R}$, we define*

- *the concave closure* $f^+(\mathbf{x}) = \max\{\sum_{S\subseteq N} \alpha_S f(S) : \sum_{S\subseteq N} \alpha_S \mathbf{1}_S = \mathbf{x}, \sum_{S\subseteq N} \alpha_S = 1, \alpha_S \geq 0\}$.
- *the convex closure* $f^-(\mathbf{x}) = \min\{\sum_{S\subseteq N} \alpha_S f(S) : \sum_{S\subseteq N} \alpha_S \mathbf{1}_S = \mathbf{x}, \sum_{S\subseteq N} \alpha_S = 1, \alpha_S \geq 0\}$.

It is easy to see by compactness that the maximum and minimum are well defined. Equivalently, we can say that $f^+(\mathbf{x}) = \max_{\mathcal{D}} \mathbb{E}_{R\sim\mathcal{D}}[f(R)]$ where the maximum is taken over all distributions $\mathcal{D}$ such that $\mathbb{E}[\mathbf{1}_R] = \mathbf{x}$. Similarly, $f^-(\mathbf{x})$ is obtained by taking the minimum over all such distributions.

**Lemma 2** *For any $f : \{0,1\}^N \to \mathbb{R}$, $f^+$ is concave and $f^-$ is convex.*

**Proof:** We prove the claim for $f^+$; the claim for $f^-$ follows by considering $-f$.

Let $\mathbf{x}, \mathbf{y} \in [0,1]^N$ and $\mathbf{z} = \lambda\mathbf{x} + (1-\lambda)\mathbf{y}$ where $\lambda \in [0,1]$. We have

$$f^+(\mathbf{x}) = \sum_{S\subseteq N} \alpha_S f(S)$$

for some probability distribution $\alpha_S$ such that $\sum \alpha_S \mathbf{1}_S = \mathbf{x}$, and similarly

$$f^+(\mathbf{y}) = \sum_{S\subseteq N} \beta_S f(S)$$

for some probability distribution $\beta_S$ such that $\sum \beta_S \mathbf{1}_S = \mathbf{y}$. Then define $\gamma_S = \lambda\alpha_S + (1-\lambda)\beta_S$, a probability distribution satisfying $\sum \gamma_S \mathbf{1}_S = \lambda\mathbf{x} + (1-\lambda)\mathbf{y} = \mathbf{z}$. By the definition of $f^+$,

$$f^+(\mathbf{z}) \geq \sum_{S\subseteq N} \gamma_S \mathbf{1}_S = \lambda f^+(\mathbf{x}) + (1-\lambda) f^+(\mathbf{y}).$$

$\square$

## 1.2 Lovász extension and submodular minimization

The convex closure for submodular function is identical to another concept - the "Lovász extension". Lovász' definition was as follows.

**Definition 3** *For a function $f : \{0,1\}^N \to \mathbb{R}$, $f^L : [0,1]^N \to \mathbb{R}$ is defined by*

$$f^L(\mathbf{x}) = \sum_{i=0}^{n} \lambda_i f(S_i)$$

*where $\emptyset = S_0 \subset S_1 \subset S_2 \subset \ldots \subset S_n$ is a chain such that $\sum \lambda_i \mathbf{1}_{S_i} = \mathbf{x}$ and $\sum \lambda_i = 1, \lambda_i \geq 0$.*

An equivalent way to define the Lovász extension is: $f^L(\mathbf{x}) = \mathbb{E}[f(\{i : x_i > \lambda\})]$, where $\lambda$ is uniformly random in $[0,1]$. Note that the Lovász extension is easy to compute, given oracle access to $f$.

**Lemma 4** *The Lovász extension $f^L$ and convex closure $f^-$ are identical if and only if $f$ is submodular.*

**Proof:** Let us assume that $f$ is submodular. Then consider $f^-(\mathbf{x}) = \min\{\sum_{S\subseteq N} \alpha_S f(S) : \sum_{S\subseteq N} \alpha_S \mathbf{1}_S = \mathbf{x}, \sum_{S\subseteq N} \alpha_S = 1, \alpha_S \geq 0\}$. Let us pick a probability distribution achieving $f^-(\mathbf{x})$ which in addition maximizes $\sum \alpha_S |S|^2$; we claim that this distribution must be supported by a chain. If not, let $\alpha_A \geq \alpha_B > 0$ be such that $A \not\subseteq B$, $B \not\subseteq A$. By submodularity, we have $f(A \cup B) + f(A \cap B) \leq f(A) + f(B)$. Let us replace an $\alpha_B$-amount of $A$ and $B$ by $A \cup B$ and $A \cap B$. This clearly does not increase $\sum \alpha_S f(S)$, and it increases $\sum \alpha_S |S|^2$:

$$|A \cup B|^2 + |A \cap B|^2 = (|A| + |B \setminus A|)^2 + (|B| - |B \setminus A|)^2 = |A|^2 + |B|^2 + 2|B \setminus A|(|A| - |B| + |B \setminus A|) > |A|^2 + |B|^2$$

since $0 \neq |B \setminus A| > |B| - |A|$. Therefore, we get a contradiction. The minimizing probability distribution $\alpha_S$ must be supported by a chain, and this is the unique chain defining $f^L(\mathbf{x})$.

If $f$ is not submodular, consider $S$ and $i, j \notin S$ such that $f(S) + f(S+i+j) > f(S+i) + f(S+j)$, and take $\mathbf{x} = \mathbf{1}_S + \frac{1}{2}\mathbf{1}_{\{i,j\}}$. The Lovász extension evaluates to

$$f^L(\mathbf{x}) = \frac{1}{2} f(S) + \frac{1}{2} f(S+i+j).$$

However, an alternative probability distribution for $\mathbf{x}$ is $\alpha_{S+i} = \alpha_{S+j} = \frac{1}{2}$, which implies

$$f^-(\mathbf{x}) \leq \frac{1}{2} f(S+i) + \frac{1}{2} f(S+j) < f^L(\mathbf{x}).$$

$\square$

This means that for submodular functions, the convex closure can be evaluated efficiently. This is not at all clear a priori, and it is not true for the concave closure which is NP-hard to evaluate for submodular functions. Since convex functions can be minimized efficiently, this explains why submodular functions can be also minimized efficiently.

**Theorem 5 (Grötschel, Lovász, Schrijver '88)** *The problem $\min_{S\subseteq N} f(S)$ can be solved in time poly$(|N|)$, for any submodular function $f : 2^N \to \mathbb{R}$.*

The GLS algorithm is based on the ellipsoid method; later, more efficient combinatorial algorihms were found by Schrijver and Fleischer-Fujishige-Iwata.

In contrast, maximizing a submodular function is NP-hard, as can be seen from the special case of Max Cut. In this sense, submodular functions are closer to convex functions than concave ones.

## 1.3 Multilinear extension

Still, submodular functions exhibit some aspects of concavity. For instance, the function $f(S) = \phi(|S|)$ is submodular if and only if $\phi$ is a concave function (of one variable). Intuitively, this concave aspect is useful in maximization problems such as $\max\{f(S) : |S| \leq k\}$ and it makes sense to look for a continuous extension of $f$ which would capture this. Unfortunately, the concave closure is hard to evaluate, so this is not the right extension for algorithmic applications. The extension which turns out to be useful here is the *multilinear extension.*

**Definition 6 (Multilinear extension)** *For a set function $f : 2^N \to \mathbb{R}$, we define its multilinear extension $F : [0,1]^N \to \mathbb{R}$ by*

$$F(\mathbf{x}) = \sum_{S \subseteq N} f(S) \prod_{i \in S} x_i \prod_{j \in N \setminus S} (1 - x_j).$$

We remark that an alternative way to define $F$ is to set $F(\mathbf{x}) = \mathbb{E}[f(\hat{\mathbf{x}})]$ where $\hat{\mathbf{x}}$ is a random set where elements appear independently with probabilities $x_i$. This is clearly equivalent to the definition above.

The multilinear extension can be defined for any set function but it acquires particularly nice properties for submodular functions.

**Lemma 7** *Let $F : [0,1]^N \to \mathbb{R}$ be the multilinear extension of a set function $f : 2^N \to \mathbb{R}$.*

- *If $f$ is non-decreasing, then $\frac{\partial F}{\partial x_i} \geq 0$ for all $i \in N$, everywhere in $[0,1]^N$.*
- *If $f$ is submodular, then $\frac{\partial^2 F}{\partial x_i \partial x_j} \leq 0$ for all $i, j \in N$, everywhere in $[0,1]^N$.*

**Proof:** Given $\mathbf{x} \in [0,1]^N$, let $R$ be a random set where elements appear independently with probabilities $x_i$. Since $F$ is multilinear, the first partial derivative $\frac{\partial F}{\partial x_i}$ is constant when only $x_i$ varies. Hence, it can be written as follows:

$$\begin{aligned}
\frac{\partial F}{\partial x_i} &= F(x_1, \ldots, x_{i-1}, 1, x_{i+1}, \ldots, x_n) - F(x_1, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_n) \\
&= \mathbb{E}[f(R + i)] - \mathbb{E}[f(R - i)] \\
&\geq 0
\end{aligned}$$

since $f(R + i) \geq f(R - i)$ by monotonicity.

To prove the second part, observe that the first partial derivatives themselves are multilinear, and hence the second partial derivatives can be written as follows:

$$\begin{aligned}
\frac{\partial^2 F}{\partial x_i \partial x_j} &= \frac{\partial F}{\partial x_j}\Big|_{(x_1, \ldots, x_{i-1}, 1, x_{i+1}, \ldots, x_n)} - \frac{\partial F}{\partial x_j}\Big|_{(x_1, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_n)} \\
&= \mathbb{E}[f(R + i + j) - f(R + i - j)] - \mathbb{E}[f(R - i + j) - f(R - i - j)] \\
&\leq 0
\end{aligned}$$

because $f(R + i + j) - f(R + i - j) \leq f(R - i + j) - f(R - i - j)$ by submodularity.

$\square$

We remark that the converse is also true: non-negativity of the first partial derivatives and non-positivity of the second partial derivatives imply monotonicity and submodularity, respectively. We leave this as an exercise for the reader. As direct consequences of Lemma 7, we obtain the following convexity properties of the multilinear relaxation.

**Corollary 8** *Let $F : [0,1]^N \to \mathbb{R}$ be the multilinear extension of a set function $f : 2^N \to \mathbb{R}$. Then*

- *If $f$ is non-decreasing, then $F$ is non-decreasing along any line of direction $\mathbf{d} \geq 0$.*
- *If $f$ is submodular, then $F$ is concave along any line of direction $\mathbf{d} \geq 0$.*
- *If $f$ is submodular, then $F$ is convex along any line of direction $\mathbf{e}_i - \mathbf{e}_j$ for $i, j \in N$.*

**Proof:** Let $\phi(\xi) = F(\mathbf{x}_0 + \xi\mathbf{d})$ denote the function along some line of direction $\mathbf{d} \geq 0$. By the chain rule and Lemma 7, we have

$$\phi'(\xi) = \sum_{i \in N} d_i \frac{\partial F}{\partial x_i}\Big|_{\mathbf{x}_0 + \xi\mathbf{d}} \geq 0$$

if $f$ is non-decreasing. In this case, $\phi$ is non-decreasing. Differentiating one more time, we get

$$\phi''(\xi) = \sum_{i,j \in N} d_i d_j \frac{\partial^2 F}{\partial x_i \partial x_j}\Big|_{\mathbf{x}_0 + \xi\mathbf{d}} \leq 0$$

by Lemma 7, if $f$ is submodular. This means that $\phi$ is concave.

Finally, consider $\psi(\xi) = F(\mathbf{x}_0 + \xi\mathbf{e}_i - \xi\mathbf{e}_j)$, the function along a line of direction $\mathbf{e}_i - \mathbf{e}_j$. Using the chain rule twice, we get

$$\psi''(\xi) = \frac{\partial^2 F}{\partial {x_i}^2} - 2\frac{\partial^2 F}{\partial x_i \partial x_j} + \frac{\partial^2 F}{\partial {x_j}^2}$$

with all the derivatives evaluated at $\mathbf{x}_0 + \xi\mathbf{e}_i - \xi\mathbf{e}_j$. We have $\frac{\partial^2 F}{\partial {x_i}^2} = \frac{\partial^2 F}{\partial {x_j}^2} = 0$ because $F$ is multilinear, and $\frac{\partial^2 F}{\partial x_i \partial x_j} \leq 0$ assuming that $f$ is submodular. Therefore, $\psi$ is convex. $\square$

## 1.4 Evaluation

There is a technical point here that we need to deal with: Evaluating the multilinear extension exactly requires $2^n$ queries to the value oracle of $f$. Obviously, this is something we cannot afford, and hence we will evaluate $F(\mathbf{x})$ only approximately.

**Lemma 9** *If $F$ is the multilinear extension of $f$, $\mathbf{x} \in [0,1]^n$, and $R_1, \ldots, R_t$ are independent samples of random sets, where element $i$ appears independently with probability $x_i$, then*

$$\left| \frac{1}{t} \sum_{i=1}^{t} f(R_i) - F(\mathbf{x}) \right| \leq \epsilon |\max f(S)|$$

*with probability at least $1 - e^{-t\epsilon^2/4}$.*

**Proof:** We can write $F(\mathbf{x}) = \mathbb{E}[f(R)]$ where $R$ is random as in the lemma. Let $M = \max |f(S)|$; $f(R)$ is a random variable in the range $[-M, M]$. Let $Y_i = \frac{1}{M} f(R_i)$ where $R_i$ is the $i$-th random sample. We have $Y_i \in [-1, 1]$ and $\sum_{i=1}^{t} \mathbb{E}[Y_i] = \frac{t}{M} F(\mathbf{x})$. By the Chernoff bound,

$$\Pr[|\sum_{i=1}^{t} Y_i - \frac{t}{M} F(\mathbf{x})| > t\epsilon] < e^{-t^2\epsilon^2/4t} = e^{-t\epsilon^2/4}.$$

$\square$

In the following, we assume that we can evaluate $F(\mathbf{x})$ to an arbitrary precision (more precisely, with additive error $|\max f(S)|/\mathrm{poly}(n)$). In some cases, further discussion is needed to ensure that this does not affect the approximation factors significantly and we shall return to this issue when necessary.

### 1.5 Summary

We have seen four (in effect three) possible extensions of a submodular function. They are ordered as follows:

$$f^+(\mathbf{x}) \geq F(\mathbf{x}) \geq f^-(\mathbf{x}) = f^L(\mathbf{x}).$$

This can be seen from the fact that each extension can be written as $\mathbb{E}[f(R)]$ for some distribution of $R$ such that $\mathbb{E}[\mathbf{1}_R] = \mathbf{x}$. The concave closure $f^+(\mathbf{x})$ maximizes this expectation, $f^-(\mathbf{x})$ minimizes the expectation, and the multilinear extension $F(\mathbf{x})$ is somewhere in between.