Parsimonious and Structured Learning
Introduction to Submodular Optimization

Uday Kiran Reddy Tadipatri [*]

July 7, 2024

This note is an introduction to submodular functions and optimization. The content is highly inspired by the monologue and slides of

In the real world, we encounter many discrete optimization problems where the feasible set is discrete, and there could be a combinatorial number of possible feasible points. In such cases, solving the optimization problem is computationally intensive, or one would have to settle for approximate solutions offered by greedy algorithms. Submodular optimization is a class of combinatorial optimization problems where the objective function is submodular. This class of problems has many interesting properties that allow us to translate these combinatorial optimization problems into convex optimization problems, enabling us to utilize tools learned in convex analysis to solve these problems efficiently.

Submodular functions arise in many applications such as machine learning, computer vision, signal processing, and electrical networks. In this note, we will introduce submodular functions, Lovász extension, and the connections between submodular and convex optimization. We shall conclude with an application to structured sparsity problems in machine learning.

1 Submodular Functions

As a notation we denote, $\mathcal{V}$ as finite set, and $2^{\mathcal{V}}$ as the power set of $\mathcal{V}$. Upper-case F denotes a function acting on the set of $2^{\mathcal{V}}$, lower-case f indicates function acting on real vectors.

Definition 1. *Submodular Function: A set-function $F : 2^{\mathcal{V}} \to \mathbb{R}$ is submodular if and only if, for all subsets $\mathcal{A}, \mathcal{B} \subseteq \mathcal{V}$, we have: $F(\mathcal{A}) + F(\mathcal{B}) \geq F(\mathcal{A} \cup \mathcal{B}) + F(\mathcal{A} \cap \mathcal{B})$.*

We always assume that, $F(\varnothing) = 0$. Note that if function, F is submodular then it implies sub-additivity, but converse is not true.

There are alternative ways to define submodular functions,

Definition 2. *First-order differences: The set-function $F : 2^{\mathcal{V}} \to \mathbb{R}$ is submodular if and only if for all subsets $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{V}$, and $i \in \mathcal{V} \backslash \mathcal{B}$, we have: $F(\mathcal{A} \cup \{i\}) - F(\mathcal{A}) \geq F(\mathcal{B} \cup \{i\}) - F(\mathcal{B})$.*

Definition 3. *Second-order differences: The set-function $F : 2^{\mathcal{V}} \to \mathbb{R}$ is submodular if and only if for all subsets $\mathcal{A} \subseteq \mathcal{V}$, and $i, j \in \mathcal{V} \backslash \mathcal{A}$, we have: $F(\mathcal{A} \cup \{i\}) - F(\mathcal{A}) \geq F(\mathcal{A} \cup \{i, j\}) - F(\mathcal{A} \cup \{j\})$.*

Verify from any of the above definition that the above function is submodular.

Example 1 (Cardinality)**.** *A simple example of submodular function is the cardinality function, $F(\mathcal{A}) = |\mathcal{A}|$. This function is also classified as modular function.*

[*]Ph.D Student, Electrical & Systems Engineering. Mail-Id: *ukreddy@seas.upenn.edu*

Example 2 (Graph Cut)**.** *Take a graph, $\mathcal{G}$ denote $\mathcal{V}(\mathcal{G}) = \{v_1, \dots, v_n\}$ to be set of vertices. The function, $F(\mathcal{S}) = |\{(u, v) : (u, v) \in \mathcal{E}(\mathcal{G}) u \in \mathcal{A}, \mathbf{v} \in \mathcal{V}(\mathcal{G}) \backslash \mathcal{A}\}|$, where, $\mathcal{S} \subseteq \mathcal{V}(\mathcal{G})$ is submodular.*

Verify the above example from First-order differences definition.

Example 3 (Mutual Information)**.** *Denote, $\Omega = \{X_1, X_2, \dots, X_n\}$ be set of random variables. The mutual information between, $\mathcal{S}$ and $\Omega \backslash \mathcal{S}$, where $\mathcal{S} \subseteq \Omega$ is submodular, i.e, $F(\mathcal{S}) = I(\mathcal{S}; \Omega \backslash \mathcal{S})$.*

Verify from the original definition that the above example is submodular.

Similar to the analogy of convex-concave function, we can also define supermodular functions. If a function is both submodular and supermodular, then it is called modular function. For instance, the earlier example is a modular function.

Now we move-on to some operations that preserve submodularity.

1. **Extension:** Let $\mathcal{B} \subseteq \mathcal{V}$, then $G : 2^{\mathcal{B}} \to \mathbb{R}$ be a submodular function, then function $F : 2^{\mathcal{V}} \to \mathbb{R}$ defined as $F(\mathcal{A}) := G(\mathcal{A} \cap \mathcal{B}); \forall \mathcal{A} \subseteq \mathcal{V}$ is also submodular.

2. **Restriction:** Let $\mathcal{B} \subseteq \mathcal{V}$, then $G : 2^{\mathcal{V}} \to \mathbb{R}$ be a submodular function, then function $F : 2^{\mathcal{B}} \to \mathbb{R}$ defined as $F(\mathcal{A}) := G(\mathcal{A}); \forall \mathcal{A} \subseteq \mathcal{B}$ is also submodular.

3. **Contraction:** Let $\mathcal{B} \subseteq \mathcal{V}$, then $G : 2^{\mathcal{V}} \to \mathbb{R}$ be a submodular function, then the function $F : 2^{\mathcal{V} \backslash \mathcal{B}} \to \mathbb{R}$ defined as $F(\mathcal{A}) := G(\mathcal{A} \cup \mathcal{B}) - G(\mathcal{B}); \forall \mathcal{A} \subseteq \mathcal{V} \backslash \mathcal{B}$ is also submodular.

4. **Partial Minimization:** Take disjoint sets $\mathcal{V}, \mathcal{W}$ and submodular function $G : 2^{\mathcal{V} \cup \mathcal{W}} \to \mathbb{R}$. Then function $F : 2^{\mathcal{V}} \to \mathbb{R}$ defined as $F(\mathcal{A}) := \min_{\mathcal{B} \subseteq \mathcal{W}} G(A \cup B) - \min_{\mathcal{B} \subseteq \mathcal{W}} G(B)$ is a submodular function.

5. **Convolution:** Let function $F : 2^{\mathcal{V}} \to \mathbb{R}$ be a submodular function, and take a vector $\mathbf{z} \in \mathbb{R}^p$. Then the fucntion $G : 2^{\mathcal{V}} \to \mathbb{R}$ defined as $G(\mathcal{A}) := \min_{\mathcal{B} \subseteq \mathcal{A}} F(\mathcal{B}) + \sum_{i \in \mathcal{A} \backslash \mathcal{B}} z_i$ is a submodular function. Moreover, $G(\mathcal{A}) \leq F(\mathcal{A}), G(\mathcal{A}) \leq \sum_{i \in \mathcal{A}} z_i$.

6. **Monotonization:** Let $F : 2^{\mathcal{V}} \to \mathbb{R}$ be a submodular function. Then function $G : 2^{\mathcal{V}} \to \mathbb{R}$ defined as $G(\mathcal{A}) := \min_{\mathcal{B} \supset \mathcal{A}} F(\mathcal{B}) - \min_{\mathcal{B} \subseteq \mathcal{A}} F(\mathcal{B})$ is submodular. Moreover, $G(\mathcal{A}) \leq F(\mathcal{A})$.

2 Lovász Extension: A bridge that connects Submodular Optimization to Convex Optimization

Definition 4. *Given a set-function $F : 2^{\mathcal{V}} \to \mathbb{R}$ (not necessarily submodular), the Lovász extension $f : \mathbb{R}^p \to \mathbb{R}$ [1] is defined as follows; for $\mathbf{w} \in \mathbb{R}^p$, order the components in decreasing order $w_{j1} \geq w_{j2} \geq \cdots \geq w_{jp}$, where $(j1, j2, \dots, jp)$ is arg sort order. The below definitions are equivalent:*

1. $f(\mathbf{w}) = \sum_{k=1}^{p} w_{jk} \left[F(\{j1, \dots jk\}) - F(\{j1, \dots jk-1\}) \right]$,

2. $f(\mathbf{w}) = \sum_{k=1}^{p-1} F(\{j1, \dots, jk\})(w_{jk} - w_{jk+1}) + F(\mathcal{V}) w_{jp}$,

3. $f(\mathbf{w}) = \int_{\min\{w_1, \dots, w_p\}}^{\infty} F(\{\mathbf{w} \geq z\}) dz + F(V) \min\{w_1, \dots, w_p\}$

4. $f(\mathbf{w}) = \int_{0}^{\infty} F(\{\mathbf{w} \geq z\}) dz + \int_{-\infty}^{0} \left[F(\{\mathbf{w} \geq z\}) - F(\mathcal{V}) \right] dz$

[1] $p = |\mathcal{V}|$

Example 4. *Let consider a set function, $F : 2^{\{1,2\}} \to \mathbb{R}$ such that, $F(\varnothing) = 0, F(\{1\}) = 1, F(\{2\}) = 1, F(\{1,2\}) = 0$.*

The Lovász extension of the above function is as follows:

$$f([x_1; x_2]) = \begin{cases} x_2 - x_1 & \text{if } x_1 \leq x_2 \\ x_1 - x_2 & \text{Otherwise} \end{cases}$$

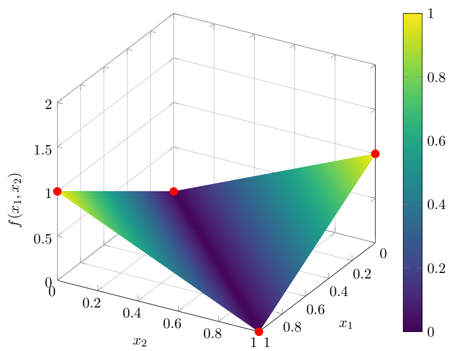

Figure 1: Lovász extension of a set-function F

Example 5. *Let consider a set-function, $F : 2^{\{1,2\}} \to \mathbb{R}$ such that, $F(\varnothing) = 0, F(\{1\}) = 1, F(\{2\}) = 1, F(\{1,2\}) = 3$.*

The Lovász extension of the above function is as follows:

$$f([x_1; x_2]) = \begin{cases} x_1 + 2x_2 & \text{if } x_1 \geq x_2 \\ 2x_1 + x_2 & \text{Otherwise} \end{cases}$$

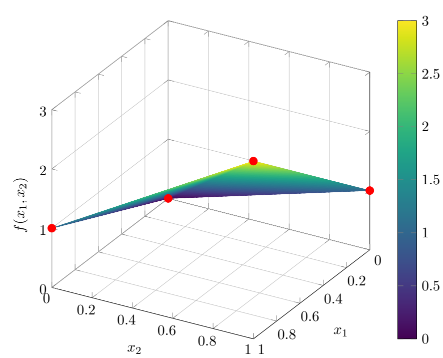

Figure 2: Lovász extension of a set-function F

Proposition 1. *Let F be any set-function, we have:*

1. *If F and G are set-functions with Lovász extensions f and g respectively, then $f + g$ is the Lovász extension of $F + G$.*
2. *For $\mathbf{w} \in \mathbb{R}^p_+$, $f(\mathbf{w}) = \int_0^\infty F(\{\mathbf{w} \geq z\})dz$.*

3. *If $F(\mathcal{V}) = 0 \implies \forall \mathbf{w} \in \mathbb{R}^p, f(\mathbf{w}) = \int_{-\infty}^{\infty} F(\{\mathbf{w} \geq z\})dz$.*

4. *For all $\mathbf{w} \in \mathbb{R}^p$ and $\alpha \in \mathbb{R}$, $f(\mathbf{w} + \alpha 1_{\mathcal{V}}) = f(\mathbf{w}) + \alpha F(\mathcal{V})$.*

5. *The Lovász extension f is positively homogeneous with degree-1.*

6. *For all $\mathcal{A} \subseteq \mathcal{V}, F(\mathcal{A}) = f(1_{\mathcal{A}})$.*

7. *If F is symmetric (i.e, $\forall \mathcal{A} \subseteq \mathcal{V}, F(\mathcal{A}) = F(\mathcal{V} \backslash \mathcal{A})$), then f is even.*

8. *If $\mathcal{V} = \mathcal{A}_1 \cup \ldots \mathcal{A}_m$ is a partition of $\mathcal{V}$, and $\mathbf{w} = \sum_{i=1}^{m} \nu_i 1_{\mathcal{A}_i}$ with $\nu_1 \geq \nu_2 \geq \ldots \nu_m$ then $f(\mathbf{w}) = \sum_{i=1}^{m-1} (\nu_i - \nu_{i+1}) F(\mathcal{A}_1 \cup \cdots \cup \mathcal{A}_i) + \nu_m F(\mathcal{V})$.*

9. *If $\mathbf{w} \in [0,1]^p$, then $f(\mathbf{w}) = \mathbb{E}_{x \sim Uniform([0,1])} [F(\{\mathbf{w} \geq x\})]$.*

Proposition 2. *A set-function F is submodular if and only if its Lovász extension f is convex.*

Proposition 3. *Let F be a submodular function and f be its Lovász extension. Then* $\min_{\mathcal{A} \subseteq \mathcal{V}} F(\mathcal{A}) = \min_{\mathbf{w} \in \{0,1\}^p} f(\mathbf{w}) = \min_{\mathbf{w} \in [0,1]^p} f(\mathbf{w})$

3 Submodular Optimization

A few properties: Let $F : 2^{\mathcal{V}} \to \mathbb{R}$ be a submodular function, then

1. **Optimality Conditions:** If $\mathcal{A} \subseteq \mathcal{V}$ is a minimizer of F if and only if $\mathcal{A}$ is minimizer of F overall subsets of $\mathcal{A}$ and all supersets of $\mathcal{A}$.

2. **Lattice of Minimizers:** If $\mathcal{A}$ and $\mathcal{B}$ are minimizers of F, then so are $\mathcal{A} \cup \mathcal{B}$ and $\mathcal{A} \cap \mathcal{B}$.

3. **Minimizing a Submodular Function:**

   Theorem 1 (Grötschel-Lovász-Schrijver '81)**.** *There is an algorithm that computes the minimum of any submodular function $F : 2^{\mathcal{V}} \to \mathbb{R}$ in $poly(|\mathcal{V}|)$ time.*

4. Maximizing a submodular function is **NP-Hard** in general.
   Greedy Algorithm: Pick elements one-by-one, maximizing the gain in $F(\mathcal{S})$, while maintaining $\mathcal{S} \subseteq \mathcal{V}$ i.e, $\mathcal{S} \leftarrow \mathcal{S} \cup \arg\max_{i \in \mathcal{S}} F(\mathcal{S} \cup \{i\}) - F(\mathcal{S})$.

   Theorem 2 (Nemhauser, Wolsey, Fisher '78)**.** *If F is non-decreasing submodular Greedy Algorithm finds a solution of value atleast $(1 - 1/e)\times$ Optimal value for the problem $\max\{F(\mathcal{S}) : |\mathcal{S}| \leq k\}$.*

   Theorem 3 (Nemhauser, Wolsey '78)**.** *No algorithm using a polynomial number of queries to F can achieve a factor better than $(1 - 1/e)$.*

3.1 Structured Sparsity

Consider the following regression problem with structured sparsity inducing norm, We can also consider the following optimization problem,

$$\min_{\mathbf{w} \in \mathbb{R}^p} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}(\mathbf{y}_i, \langle \mathbf{w}, \mathbf{x}_i \rangle) + \lambda' F(S(\mathbf{w}))$$

where, here, S is the support of $\mathbf{w}$, F is set-function penalizes discretely over the support set (it is submodular). The above problem is a combinatorial optimization problem, and is computationally intensive. We can relax the above to below problem,

$$\min_{\mathbf{w}\in\mathbb{R}^p} \frac{1}{N}\sum_{i=1}^{N} \mathcal{L}(\mathbf{y}_i, \langle \mathbf{w}, \mathbf{x}_i\rangle) + \lambda\Omega_S(\mathbf{w})$$

where, Ω_S is a gauge function, and λ is a regularization parameter. The above problem is a convex optimization problem.

We ideally want a optimal convex relaxation of the original problem. The below proposition gives us a way to do that.

Proposition 4. *Let F be a non-decreasing submodular function. The function $\Omega_S : \mathbf{w} \to f(|\mathbf{w}|)$ is the tightest convex lower bound (dual of dual) of the function $\mathbf{w} \to F(S(\mathbf{w}))$ on the unit L_∞ ball, $[-1,1]^p$, here f is the Lovász extension of F.*

LASSO: The Lovász extension of the cardinality function, $F(\mathcal{A}) := |\mathcal{A}|$ is $f(\mathbf{w}) = \langle \mathbf{w}, 1_{\mathcal{V}}\rangle$. Therefore, the tightest convex lower bound of L_0 norm is L_1 norm.

Let consider a set-function, $F : 2^{\{1,2\}} \to \mathbb{R}$ such that, $F(\varnothing) = 0, F(\{1\}) = 1, F(\{2\}) = 1, F(\{1,2\}) = 2$.

We have that $f([|x_1|;|x_2|])$

$$f([x_1;x_2]) = |x_1| + |x_2|$$

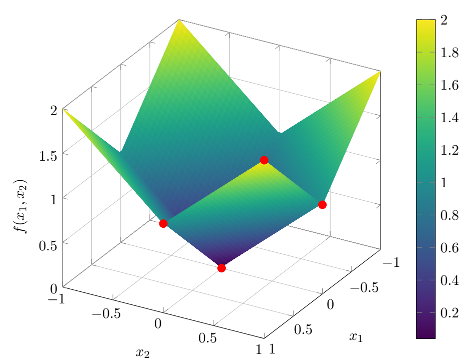

Figure 3: Lovász extension of a set-function F