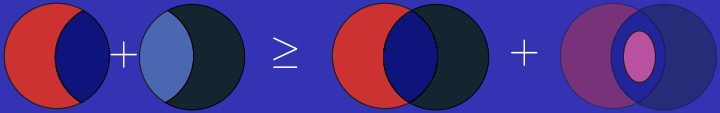


Submodularity in Machine Learning

Submodularity, Convexity and Concavity

Rishabh Iyer and Jeffrey A. Bilmes

Departments of Electrical Engineering
& Computer Science and Engineering
University of Washington, Seattle

June 18th, 2014

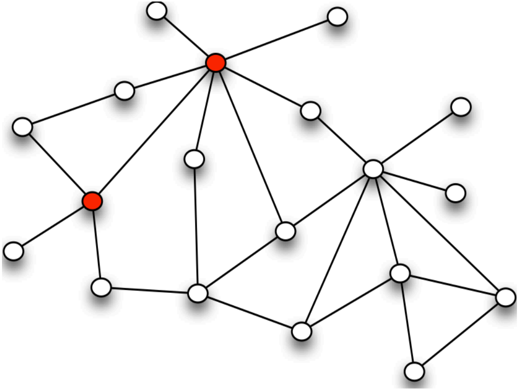
Outline

Acknowledgments

Special Thanks to my colleagues and collaborators, Jeff Bilmes, Stefanie Jegelka, Kai Wei, Jennifer Gillenwater, Ganesh Ramakrishnan, Sebastian Tschiatschek, Yoshinobu Kawahara, Ganesh Ramakrishnan, Matthai Phillipose, Bethany Herwaldt, Shengjie Wang, Wenruo Bai.

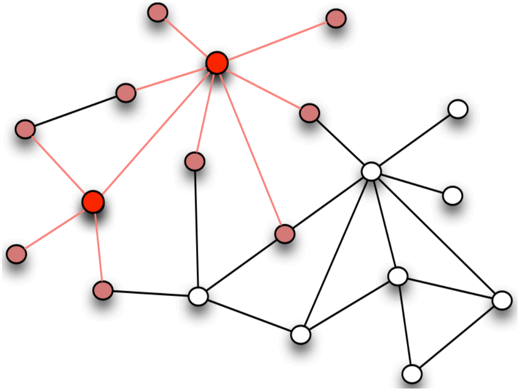
Outline

Recap: Set functions



Modeling Social Influence

Recap: Set functions



Modeling Social Influence

Recap: Set functions

Representative Sentences

Recap: Set functions

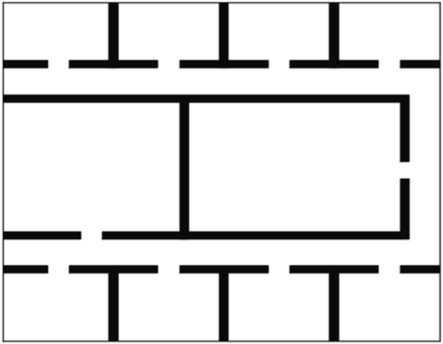
Representative Sentences

Recap: Set functions

Summarizing Image Collections?

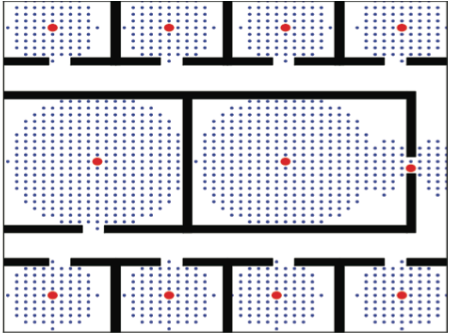
Recap: Set functions

Recap: Set functions



Sensor placement

Recap: Set functions



Sensor placement

Recap: Set functions

Segmenting Images

Recap: Set functions

What's common?

Optimize a set function f(A) under constraints!

Recap: Set functions

$$V = \{ \quad \}$$

$$f : 2^V \to \mathbb{R}$$

Recap: Set functions

$V =$

Choose Subset

$$A \subseteq V$$

Recap: Set functions

$A = \{$ , , , , $\}$

$$f(A) = 22$$

Recap: Set functions

$$f : 2^V \to \mathbb{R}$$

General Set function Optimization –
very very hard!

What if there is some special structure?

Recap: Submodular Functions

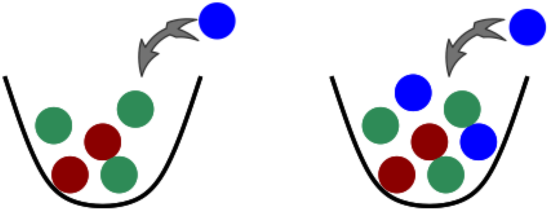
- Special class of set functions.

$$f(A \cup v) - f(A) \geq f(B \cup v) - f(B), \text{ if } A \subseteq B \quad (1)$$

Recap: Submodular Functions

- Special class of set functions.

$$f(A \cup v) - f(A) \geq f(B \cup v) - f(B), \text{ if } A \subseteq B \tag{1}$$

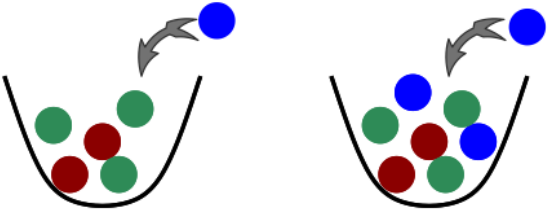


$f = \#$ of distinct colors of balls in the urn.

Recap: Submodular Functions

- Special class of set functions.

$$f(A \cup v) - f(A) \geq f(B \cup v) - f(B), \text{ if } A \subseteq B \quad (1)$$



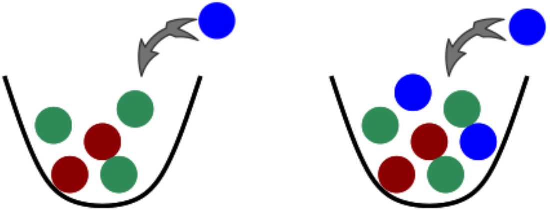
$f = \#$ of distinct colors of balls in the urn.

Gain $= 1$

Recap: Submodular Functions

- Special class of set functions.

$$f(A \cup v) - f(A) \geq f(B \cup v) - f(B), \text{ if } A \subseteq B \quad (1)$$



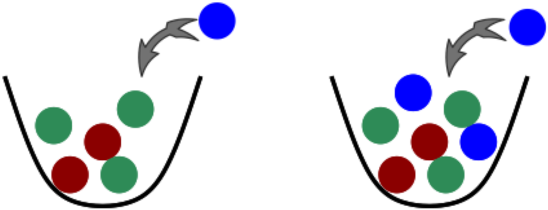
$f = \#$ of distinct colors of balls in the urn.

Gain $= 1$ Gain $= 0$

Recap: Submodular Functions

- Special class of set functions.

$$f(A \cup v) - f(A) \geq f(B \cup v) - f(B), \text{ if } A \subseteq B \qquad (1)$$



$f = \#$ of distinct colors of balls in the urn.

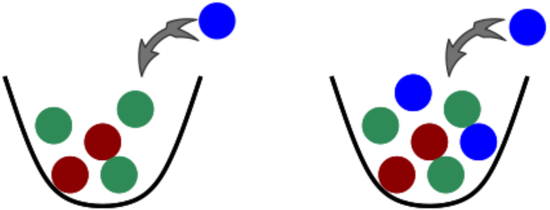
Gain $= 1$ Gain $= 0$

- Monotonicity: $\mathbf{f(A) \leq f(B)}$, if $\mathbf{A \subseteq B}$.

Recap: Submodular Functions

- Special class of set functions.

$$f(A \cup v) - f(A) \geq f(B \cup v) - f(B), \text{ if } A \subseteq B \quad (1)$$



$f = \#$ of distinct colors of balls in the urn.

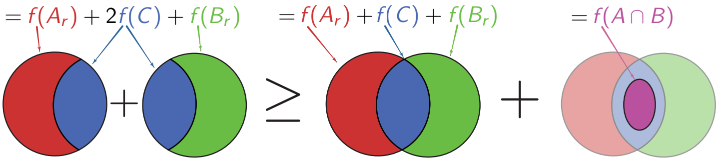
Gain $= 1$ Gain $= 0$

- Monotonicity: $\mathbf{f}(\mathbf{A}) \leq \mathbf{f}(\mathbf{B})$, if $\mathbf{A} \subseteq \mathbf{B}$.
- Modular function $\mathbf{f}(\mathbf{X}) = \sum_{\mathbf{i} \in \mathbf{X}} \mathbf{f}(\mathbf{i})$ analogous to linear functions.

Recap: Alternate definition – Submodular Functions

- A function $f : 2^V \rightarrow \mathbb{R}$ is submodular if:

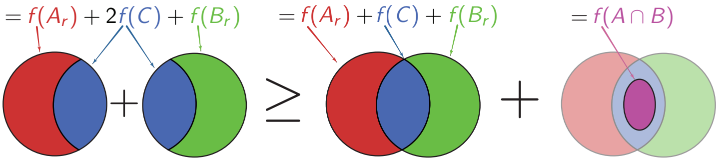
$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B)$$



Recap: Alternate definition – Submodular Functions

- A function $f : 2^V \to \mathbb{R}$ is submodular if:

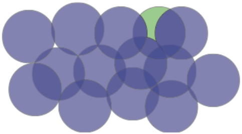
$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B)$$



- Submodularity has been widely used: non-additive measure theory, economics, game theory, statistical physics and thermodynamics, electrical networks, and operations research.

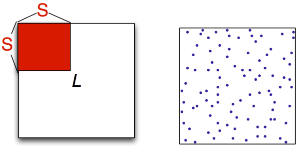
Facets of Submodular functions (models in maximization)

$$F(A) = \cup_{s \in A} \text{area}(s)$$



Coverage

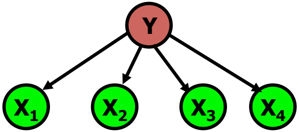
$$F(A) = \log det(L_A)$$



$L \in \mathbb{R}^{p \times p}$ psd

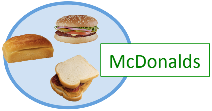
Diversity

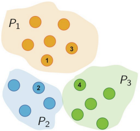
$$F(A) = H(X_A)$$



Information

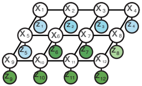
Facets of Submodular functions (models in minimization)

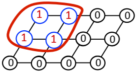




$$F(A) = \sum_{i=1}^{3} \sqrt{\sum_{j \in A \cap P_i} r_i}$$

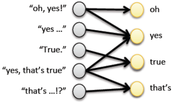
Cooperative Costs



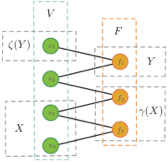


$$E(\mathbf{x};\mathbf{z}) = \sum_i E_i(x_i) + \sum_{ij} E_{ij}(x_i, x_j)$$

Attractive Potentials



$$F(A) = \gamma(A)$$



(a) Bipartite graph

Complexity

Overview of this part of the tutorial

- Submodularity, Convexity and Concavity.
- Polyhedra associated with submodular functions.
- Submodular Semigradients.
- Convex and Concave extensions of Submodular Functions.

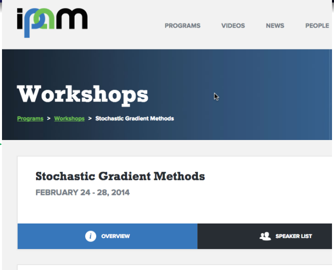
Next Part of this tutorial

- Unifying class of algorithms for submodular minimization, submodular maximization, DS optimization and submodular optimization subject to submodular constraints.
- Most of these algorithms are based on the convex and concave aspects of submodelar functions.
- Extensions to Submodular Structures including Submodular partitioning, Submodular metrics, Submodular Bregman and Submodular Point Processes.

Outline

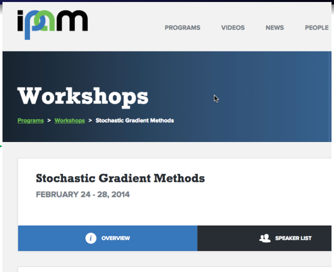
1. Introduction
2. Submodularity, Convexity and Concavity
3. Submodularity and Convexity
   - Submodular Polyhedron
   - Convex extension
   - Submodular Subdifferential
   - Generalized lower Submodular Polyhedron
   - Convex aspects of a submodular function
4. Submodularity and Concavity
   - Submodular Upper Polyheron
   - Submodular Superdifferentials
   - Generalized Upper Polyhedron
   - Continuous extensions of a submodular function
   - Concave Aspects of a Submodular Function

Convexity and Gradients



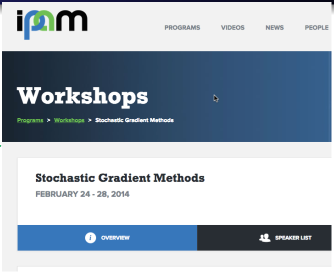
- Big training data in machine learning: computational biology, speech and language processing, collaborative filtering, computer vision.

Convexity and Gradients



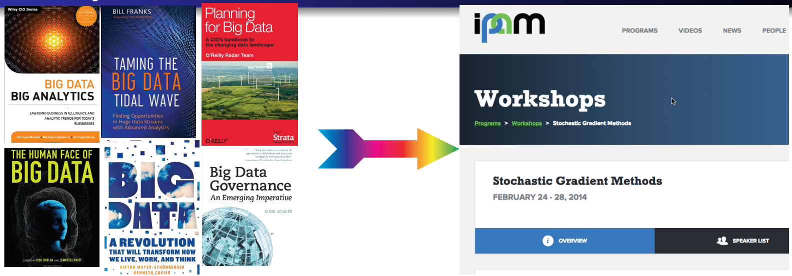
- Big training data in machine learning: computational biology, speech and language processing, collaborative filtering, computer vision.
- Motivates stochastic approximation/stochastic gradient methods, often effective on large scale machine learning problems.

Convexity and Gradients



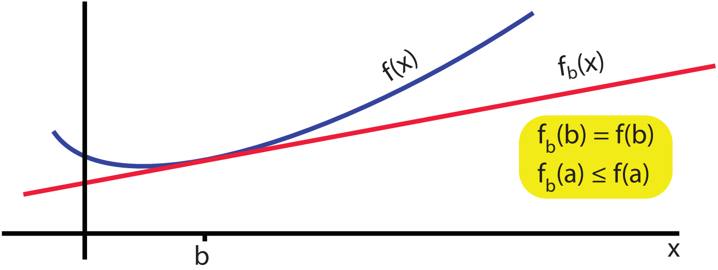
- Big training data in machine learning: computational biology, speech and language processing, collaborative filtering, computer vision.
- Motivates stochastic approximation/stochastic gradient methods, often effective on large scale machine learning problems.
- Readily applied now both to convex and non-convex problems.

Convexity and Gradients



- Big training data in machine learning: computational biology, speech and language processing, collaborative filtering, computer vision.
- Motivates stochastic approximation/stochastic gradient methods, often effective on large scale machine learning problems.
- Readily applied now both to convex and non-convex problems.
- Some Methods: Conditional Gradient, Subgradient/Mirror Descent, Generalized Accelerated Gradient Ascent (GAGA), Incremental Gradient, Nesterov's Optimal Gradient, Proximal Gradient, Fast Proximal Gradient, etc.

Convex Functions and Tight Subgradients

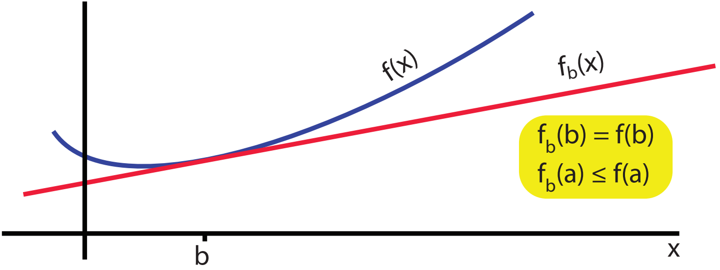


- A convex function f has a subgradient at any in-domain point b, namely there exists f_b such that

$$f(x) - f(b) \geq \langle f_b, x - b \rangle, \forall x. \quad (2)$$

we have $f_b(x) = f(b) + \langle f_b, x - b \rangle$

Convex Functions and Tight Subgradients



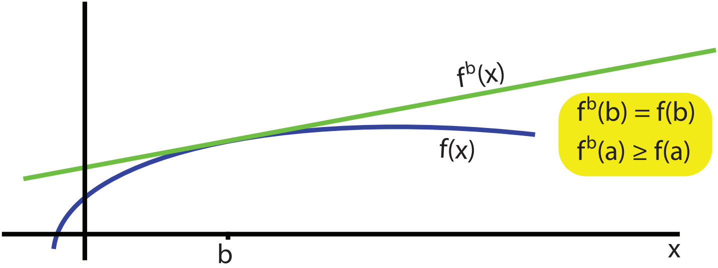
- A convex function f has a subgradient at any in-domain point b, namely there exists f_b such that

$$f(x) - f(b) \geq \langle f_b, x - b \rangle, \forall x. \tag{2}$$

we have $f_b(x) = f(b) + \langle f_b, x - b \rangle$

- We have that $f(x)$ is convex, $f_b(x)$ is affine, and can be a tight subgradient (tight at b, affine lower bound on $f(x)$) for all b

Concave Functions and Tight Supergradients

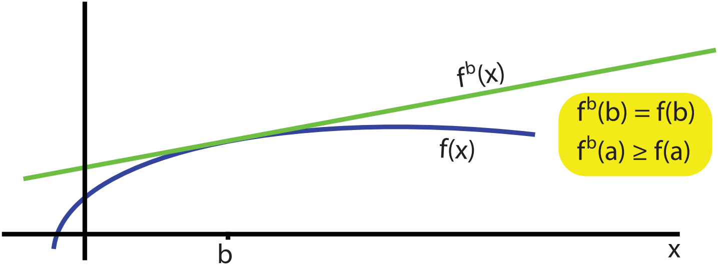


- A concave f has a supergradient at any in-domain point b, namely there exists f^b such that

$$f(x) - f(b) \leq \langle f^b, x - b \rangle, \forall x. \tag{3}$$

we have $f^b(x) = f(b) + \langle f^b, x - b \rangle$

Concave Functions and Tight Supergradients



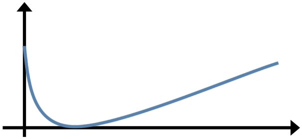
- A concave f has a supergradient at any in-domain point b, namely there exists f^b such that

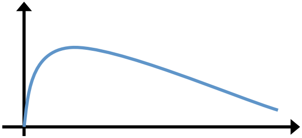
$$f(x) - f(b) \leq \langle f^b, x - b \rangle, \forall x. \tag{3}$$

  we have $f^b(x) = f(b) + \langle f^b, x - b \rangle$
- We have that $f(x)$ is concave, $f^b(x)$ is affine, and can be a tight supergradient (tight at b, affine upper bound on $f(x)$) for all b

Two Sides of Submodularity

discrete convexity ....





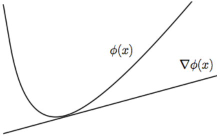
... or concavity?

Submodular functions have properties related to both!

Two sides of Submodularity

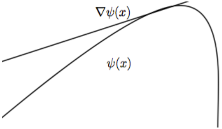
Convex aspects (Fujishige (1984, 2005), Frank (1982))

- Minimization: Poly-time.
- Convex continuous extension - Lovász extension.
- Subgradients and Subdifferential.
- Convex duality, discrete seperation etc.



Concave aspects (Vondrak (2007), I-Bilmes (2015))

- Max: constant-factor approx!
- Multilinear extension - concave in a direction.
- Supergradients and Superdifferential.
- Under restricted settings, duality, separation etc.



Submodular Semigradients and extensions

- A submodular function $f : 2^V \to \mathbb{R}$, has both tight subgradients and supergradients, tight at a set $Y \subseteq V$:
  1. Tight Subgradients: $\exists m_Y \in \mathbb{R}^V$ and $b_Y \in \mathbb{R}$ such that $m_Y(Y) + b_Y = f(Y)$ and $m_Y(X) + b_Y \leq f(X)$ for all $X \subseteq V$.

Submodular Semigradients and extensions

- A submodular function $f : 2^V \rightarrow \mathbb{R}$, has both tight subgradients and supergradients, tight at a set $Y \subseteq V$:
  1. Tight Subgradients: $\exists m_Y \in \mathbb{R}^V$ and $b_Y \in \mathbb{R}$ such that $m_Y(Y) + b_Y = f(Y)$ and $m_Y(X) + b_Y \leq f(X)$ for all $X \subseteq V$.
  2. Tight Supergradients: $\exists m^Y \in \mathbb{R}^V$ and $b^Y \in \mathbb{R}$ such that $m^Y(Y) + b^Y = f(Y)$ and $m^Y(X) + b^Y \leq f(X)$ for all $X \subseteq V$.

Submodular Semigradients and extensions

- A submodular function $f : 2^V \to \mathbb{R}$, has both tight subgradients and supergradients, tight at a set $Y \subseteq V$:
  1. Tight Subgradients: $\exists m_Y \in \mathbb{R}^V$ and $b_Y \in \mathbb{R}$ such that $m_Y(Y) + b_Y = f(Y)$ and $m_Y(X) + b_Y \leq f(X)$ for all $X \subseteq V$.
  2. Tight Supergradients: $\exists m^Y \in \mathbb{R}^V$ and $b^Y \in \mathbb{R}$ such that $m^Y(Y) + b^Y = f(Y)$ and $m^Y(X) + b^Y \leq f(X)$ for all $X \subseteq V$.

Submodular Semigradients and extensions

- A submodular function $f : 2^V \to \mathbb{R}$, has both tight subgradients and supergradients, tight at a set $Y \subseteq V$:
  1. Tight Subgradients: $\exists m_Y \in \mathbb{R}^V$ and $b_Y \in \mathbb{R}$ such that $m_Y(Y) + b_Y = f(Y)$ and $m_Y(X) + b_Y \leq f(X)$ for all $X \subseteq V$.
  2. Tight Supergradients: $\exists m^Y \in \mathbb{R}^V$ and $b^Y \in \mathbb{R}$ such that $m^Y(Y) + b^Y = f(Y)$ and $m^Y(X) + b^Y \leq f(X)$ for all $X \subseteq V$.
- Submodular functions also admit continuous extensions which are convex, concave, and multilinear.

Outline

Outline

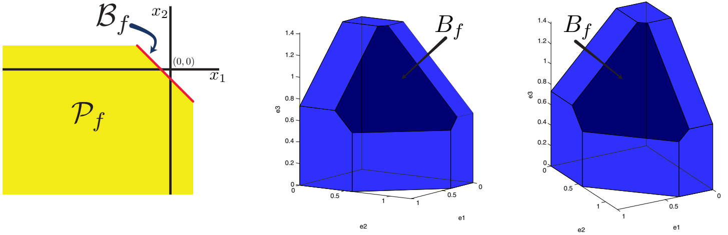
Submodular (lower) Polyhedron

- A submodular $f : 2^V \to \mathbb{R}$, has a polyhedron called the submodular (lower) polyhedron and a base (lower) polytope:

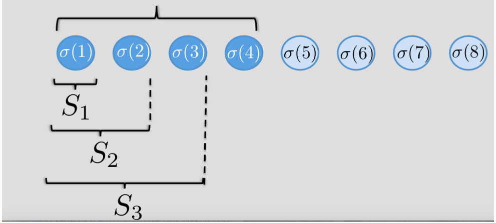
$$\mathcal{P}_f = \{x \in \mathbb{R}^V : x(S) \leq f(S), \forall S \subseteq V\} \quad (4)$$

$$\mathcal{B}_f = \mathcal{P}_f \cap \{x : x(V) = f(V)\}. \quad (5)$$

where $x(S) = \sum_{i \in S} x_i$ is seen as a modular function



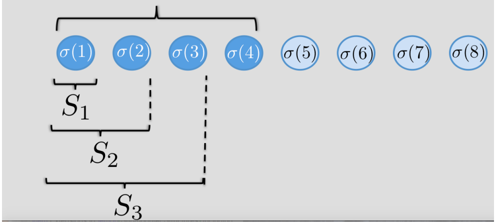
Chains & Extreme Points of the Submodular Polyhedron



- Notation: Given a permutation $\sigma = (\sigma(1), \sigma(2), \ldots, \sigma(n))$ of V, define chain $\emptyset = S_0^\sigma \subset S_1^\sigma \subset \cdots \subset S_n^\sigma = V$ where

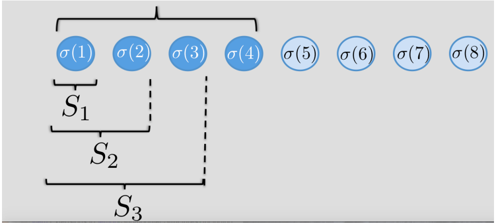
$$S_i^\sigma = \{\sigma(1), \sigma(2), \ldots, \sigma(i)\} \quad (6)$$

Chains & Extreme Points of the Submodular Polyhedron



- Notation: Given a permutation $\sigma = (\sigma(1), \sigma(2), \ldots, \sigma(n))$ of V, define chain $\emptyset = S_0^\sigma \subset S_1^\sigma \subset \cdots \subset S_n^\sigma = V$ where
$$S_i^\sigma = \{\sigma(1), \sigma(2), \ldots, \sigma(i)\} \tag{6}$$
- These chains define all extreme points.

Chains & Extreme Points of the Submodular Polyhedron

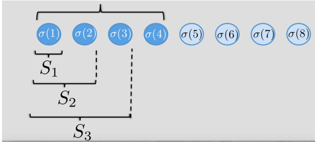


- Notation: Given a permutation $\sigma = (\sigma(1), \sigma(2), \ldots, \sigma(n))$ of V, define chain $\emptyset = S_0^\sigma \subset S_1^\sigma \subset \cdots \subset S_n^\sigma = V$ where
$$S_i^\sigma = \{\sigma(1), \sigma(2), \ldots, \sigma(i)\} \tag{6}$$
- These chains define all extreme points.
- (Edmonds, 1970) Define $h^\sigma \in \mathbb{R}^V$ as,
$$h^\sigma(\sigma(i)) = f(S_i^\sigma) - f(S_{i-1}^\sigma) \tag{7}$$
Then, h^σ is an extreme point of $\mathcal{P}_f$.

Outline

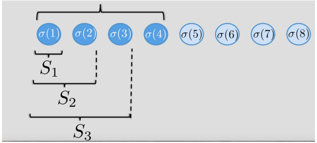
Lovász extension of a submodular function

- Given the submodular polyhedron, we can define a convex extension of a submodular function as, $\breve{f}(w) = \max_{s \in \mathcal{P}_f} w^\intercal s$.



Lovász extension of a submodular function

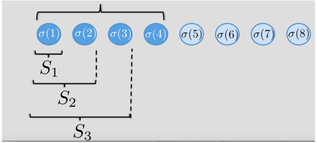
- Given the submodular polyhedron, we can define a convex extension of a submodular function as, $\breve{f}(w) = \max_{s \in \mathcal{P}_f} w^\intercal s$.



- Given vector $w \in \mathbb{R}_+^V$, define w-cognizant permutation σ_w such that $w[\sigma_w(1)] \geq w[\sigma_w(2)] \geq \cdots \geq w[\sigma_w(n)]$.

Lovász extension of a submodular function

- Given the submodular polyhedron, we can define a convex extension of a submodular function as, $\breve{f}(w) = \max_{s \in \mathcal{P}_f} w^\intercal s$.



- Given vector $w \in \mathbb{R}_+^V$, define w-cognizant permutation σ_w such that $w[\sigma_w(1)] \geq w[\sigma_w(2)] \geq \cdots \geq w[\sigma_w(n)]$.
- (Vitali'25, Choquet'54, Edmonds'70, Lovász'83): The Lovász extension is,

$$\breve{f}(w) \triangleq \max_{s \in P_f} w^\top s = \sum_{i=1}^{n} w(\sigma_w(i))[f(S_i^{\sigma_w}) - f(S_{i-1}^{\sigma_w})] = \sum_{i=1}^{n} \lambda_i f(S_i^{\sigma_w})$$

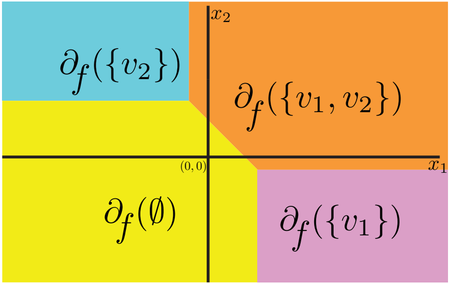
with $\sum_i \lambda_i \mathbf{1}_{S_i^{\sigma_w}} = w, \sum_i \lambda_i = 1$.

Outline

Submodular Subdifferential

- Analogous to convex functions, submodular functions have subdifferential structure (Fujishige'84,'05) at each $X \subseteq V$.

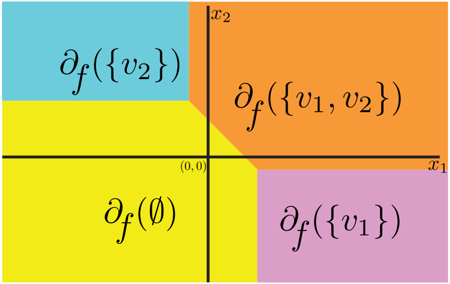
$$\partial_f(X) = \{x \in \mathbb{R}^n : f(Y) - x(Y) \geq f(X) - x(X) \; \forall Y \subseteq V\} \quad (8)$$



Submodular Subdifferential

- Analogous to convex functions, submodular functions have subdifferential structure (Fujishige'84,'05) at each $X \subseteq V$.

$$\partial_f(X) = \{x \in \mathbb{R}^n : f(Y) - x(Y) \geq f(X) - x(X) \; \forall Y \subseteq V\} \quad (8)$$



- Each $h_X \in \partial_f(X)$ defines modular lower bound of f tight at X:

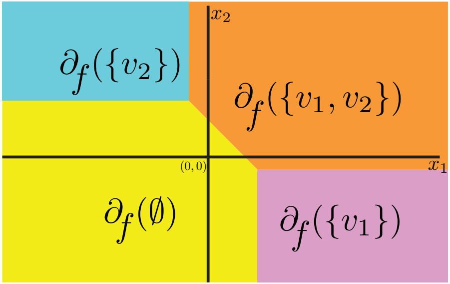
$$\partial_f(X) = \{x \in \mathbb{R}^n : f(Y) \geq f(X) - x(X) + x(Y); \forall Y \subseteq V\}$$

so $m_X(Y) \triangleq f(X) - h_X(X) + h_X(Y) \leq f(Y)$ and $m_X(X) = f(X)$.

Submodular Subdifferential

- Analogous to convex functions, submodular functions have subdifferential structure (Fujishige'84,'05) at each $X \subseteq V$.

$$\partial_f(X) = \{x \in \mathbb{R}^n : f(Y) - x(Y) \geq f(X) - x(X) \; \forall Y \subseteq V\} \quad (8)$$



- Each $h_X \in \partial_f(X)$ defines modular lower bound of f tight at X:

$$\partial_f(X) = \{x \in \mathbb{R}^n : f(Y) \geq f(X) - x(X) + x(Y); \forall Y \subseteq V\}$$

so $m_X(Y) \triangleq f(X) - h_X(X) + h_X(Y) \leq f(Y)$ and $m_X(X) = f(X)$.

- m_X is a (not necessarily normalized) modular function.

Submodular Subdifferential Extreme Points

- Given permutation σ with $S^\sigma_{|X|} = X$, take $h^\sigma_X \in \mathbb{R}^V$ with entries $h^\sigma_X(\sigma(i)) = f(S^\sigma_i) - f(S^\sigma_{i-1})$.

Submodular Subdifferential Extreme Points

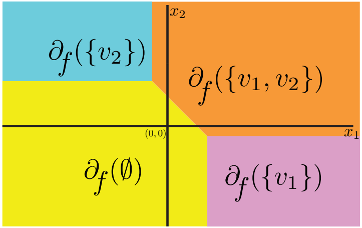
- Given permutation σ with $S^\sigma_{|X|} = X$, take $h^\sigma_X \in \mathbb{R}^V$ with entries $h^\sigma_X(\sigma(i)) = f(S^\sigma_i) - f(S^\sigma_{i-1})$.
- Then $h^\sigma_X \in \partial_f(X)$ and is itself a normalized modular function with $h_X(Y) \leq f(Y) \forall Y$ and $h_X(X) = f(X)$.

Submodular Subdifferential Extreme Points

- Given permutation σ with $S^\sigma_{|X|} = X$, take $h^\sigma_X \in \mathbb{R}^V$ with entries $h^\sigma_X(\sigma(i)) = f(S^\sigma_i) - f(S^\sigma_{i-1})$.
- Then $h^\sigma_X \in \partial_f(X)$ and is itself a normalized modular function with $h_X(Y) \leq f(Y) \forall Y$ and $h_X(X) = f(X)$.
- Such h^σ_X, for the various orders σ with property $S^\sigma_{|X|} = X$, comprise the extreme points of $\partial_f(X)$ (Fujishige,'05).

Submodular Subdifferential Extreme Points

- Given permutation σ with $S^\sigma_{|X|} = X$, take $h^\sigma_X \in \mathbb{R}^V$ with entries $h^\sigma_X(\sigma(i)) = f(S^\sigma_i) - f(S^\sigma_{i-1})$.
- Then $h^\sigma_X \in \partial_f(X)$ and is itself a normalized modular function with $h_X(Y) \leq f(Y) \forall Y$ and $h_X(X) = f(X)$.
- Such h^σ_X, for the various orders σ with property $S^\sigma_{|X|} = X$, comprise the extreme points of $\partial_f(X)$ (Fujishige,'05).
- Hence, $\partial_f(\emptyset)$ and $\partial_f(V)$ have the same set of extreme points, and $\partial_f(Y)$ for $\emptyset \subset Y \subset V$ have fewer.



Submodular Subdifferential Redundancy

- Define three polyhedra based on a partition of the constraints:

$$\partial^1_f(X) = \{x \in \mathbb{R}^n : f(Y) - x(Y) \geq f(X) - x(X), \forall Y \subseteq X\} \quad (9)$$

$$\partial^2_f(X) = \{x \in \mathbb{R}^n : f(Y) - x(Y) \geq f(X) - x(X), \forall Y \supseteq X\} \quad (10)$$

$$\partial^3_f(X) = \{x \in \mathbb{R}^n : f(Y) - x(Y) \geq f(X) - x(X), \forall Y : Y \not\subseteq X, Y \not\supseteq X\} \quad (11)$$

Submodular Subdifferential Redundancy

- Define three polyhedra based on a partition of the constraints:

$$\partial_f^1(X) = \{x \in \mathbb{R}^n : f(Y) - x(Y) \geq f(X) - x(X), \forall Y \subseteq X\} \quad (9)$$

$$\partial_f^2(X) = \{x \in \mathbb{R}^n : f(Y) - x(Y) \geq f(X) - x(X), \forall Y \supseteq X\} \quad (10)$$

$$\partial_f^3(X) = \{x \in \mathbb{R}^n : f(Y) - x(Y) \geq f(X) - x(X), \forall Y : Y \not\subseteq X, Y \not\supseteq X\} \quad (11)$$

- Immediately $\partial_f(X) = \partial_f^1(X) \cap \partial_f^2(X) \cap \partial_f^3(X)$ but more interestingly:

Submodular Subdifferential Redundancy

- Define three polyhedra based on a partition of the constraints:

$$\partial_f^1(X) = \{x \in \mathbb{R}^n : f(Y) - x(Y) \geq f(X) - x(X), \forall Y \subseteq X\} \tag{9}$$

$$\partial_f^2(X) = \{x \in \mathbb{R}^n : f(Y) - x(Y) \geq f(X) - x(X), \forall Y \supseteq X\} \tag{10}$$

$$\partial_f^3(X) = \{x \in \mathbb{R}^n : f(Y) - x(Y) \geq f(X) - x(X), \forall Y : Y \not\subseteq X, Y \not\supseteq X\} \tag{11}$$

- Immediately $\partial_f(X) = \partial_f^1(X) \cap \partial_f^2(X) \cap \partial_f^3(X)$ but more interestingly:

Lemma 1 (Fujishige'84)

Given a submodular function, $\partial_f(X) = \partial_f^1(X) \cap \partial_f^2(X)$ for all $X \subseteq V$.

Submodular Subdifferential Redundancy

- Define three polyhedra based on a partition of the constraints:

$$\partial_f^1(X) = \{x \in \mathbb{R}^n : f(Y) - x(Y) \geq f(X) - x(X), \forall Y \subseteq X\} \quad (9)$$

$$\partial_f^2(X) = \{x \in \mathbb{R}^n : f(Y) - x(Y) \geq f(X) - x(X), \forall Y \supseteq X\} \quad (10)$$

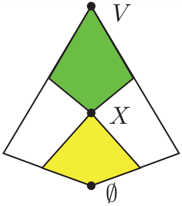
$$\partial_f^3(X) = \{x \in \mathbb{R}^n : f(Y) - x(Y) \geq f(X) - x(X), \forall Y : Y \not\subseteq X, Y \not\supseteq X\} \quad (11)$$

- Immediately $\partial_f(X) = \partial_f^1(X) \cap \partial_f^2(X) \cap \partial_f^3(X)$ but more interestingly:

Lemma 1 (Fujishige'84)

Given a submodular function, $\partial_f(X) = \partial_f^1(X) \cap \partial_f^2(X)$ for all $X \subseteq V$.

- So for $X \notin \{\emptyset, V\}$ many of the subdifferential inequalities are redundant.



Submodular Subdifferential Outer Bound

- Hence, we can write the subdifferential as:

$$\partial_f(X) = \{\{x \in \mathbb{R}^n : f(Y) - x(Y) \geq f(X) - x(X) \; \forall Y \in [\emptyset, X] \cup [X, V]\}$$

where $[A, B] = \{X \subseteq V : A \subseteq X \subseteq B\}$ whenever $A \subseteq B$.

Submodular Subdifferential Outer Bound

- Hence, we can write the subdifferential as:

$$\partial_f(X) = \{\{x \in \mathbb{R}^n : f(Y) - x(Y) \geq f(X) - x(X) \; \forall Y \in [\emptyset, X] \cup [X, V]\}$$

where $[A, B] = \{X \subseteq V : A \subseteq X \subseteq B\}$ whenever $A \subseteq B$.

- Also, submodular polyhedron $\mathcal{P}_f = \partial_f(\emptyset)$.

Submodular Subdifferential Outer Bound

- Hence, we can write the subdifferential as:

$$\partial_f(X) = \{\{x \in \mathbb{R}^n : f(Y) - x(Y) \geq f(X) - x(X) \; \forall Y \in [\emptyset, X] \cup [X, V]\}$$

where $[A, B] = \{X \subseteq V : A \subseteq X \subseteq B\}$ whenever $A \subseteq B$.

- Also, submodular polyhedron $\mathcal{P}_f = \partial_f(\emptyset)$.
- Consider polyhedron defined only for Y such that $|Y \setminus X| = |X \setminus Y| = 1$:

$$\partial_f^{\Delta(1,1)}(X) = \{x \in \mathbb{R}^V : \forall j \in X, f(j|X \backslash j) \leq x(j) \text{ and } \forall j \notin X, f(j|X) \geq x(j)\}. \tag{12}$$

where $f(j|A) = f(A + j) - f(A)$.

Submodular Subdifferential Outer Bound

- Hence, we can write the subdifferential as:

$$\partial_f(X) = \{\{x \in \mathbb{R}^n : f(Y) - x(Y) \geq f(X) - x(X) \ \forall Y \in [\emptyset, X] \cup [X, V]\}$$

  where $[A, B] = \{X \subseteq V : A \subseteq X \subseteq B\}$ whenever $A \subseteq B$.
- Also, submodular polyhedron $\mathcal{P}_f = \partial_f(\emptyset)$.
- Consider polyhedron defined only for Y such that $|Y \setminus X| = |X \setminus Y| = 1$:

$$\partial_f^{\Delta(1,1)}(X) = \{x \in \mathbb{R}^V : \forall j \in X, f(j|X \backslash j) \leq x(j) \text{ and } \forall j \notin X, f(j|X) \geq x(j)\}. \tag{12}$$

  where $f(j|A) = f(A + j) - f(A)$.
- Immediately, $\partial_f^{\Delta(1,1)}(X) \supseteq \partial_f(X)$.

Outline

Generalized (lower) Polyhedron

- Define a generalization of the submodular polyhedron $\mathcal{P}_f^{\text{gen}} \subseteq \mathbb{R}^{|V|+1}$:

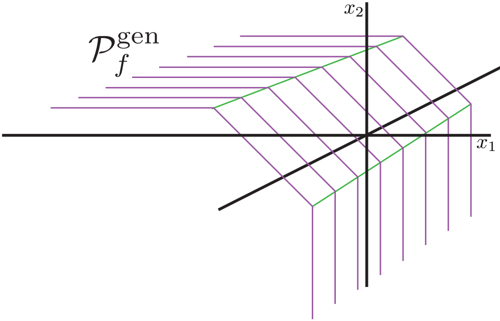
$$\mathcal{P}_f^{\text{gen}} = \{(x, c), x \in \mathbb{R}^n, c \in \mathbb{R} : [x(X) + c] \leq f(X), \forall X \subseteq V\} \quad (13)$$

Generalized (lower) Polyhedron

- Define a generalization of the submodular polyhedron $\mathcal{P}_f^{\text{gen}} \subseteq \mathbb{R}^{|V|+1}$:

$$\mathcal{P}_f^{\text{gen}} = \{(x, c), x \in \mathbb{R}^n, c \in \mathbb{R} : [x(X) + c] \leq f(X), \forall X \subseteq V\} \quad (13)$$

- For normalized submodular functions ($f(\emptyset) = 0$), we have $c \leq 0$, so extreme points exist, yielding Generalized Submodular (lower) Polyhedron:

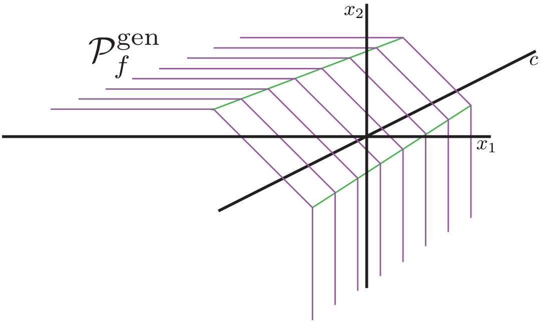


Generalized (lower) Polyhedron

- Define a generalization of the submodular polyhedron $\mathcal{P}_f^{\text{gen}} \subseteq \mathbb{R}^{|V|+1}$:

$$\mathcal{P}_f^{\text{gen}} = \{(x, c), x \in \mathbb{R}^n, c \in \mathbb{R} : [x(X) + c] \leq f(X), \forall X \subseteq V\} \quad (13)$$

- For normalized submodular functions ($f(\emptyset) = 0$), we have $c \leq 0$, so extreme points exist, yielding Generalized Submodular (lower) Polyhedron:



- Immediately, $\mathcal{P}_f \times \{0\} = \{(x, c) \in \mathbb{R}^{|V|+1} : c = 0\} \cap \mathcal{P}_f^{\text{gen}}$.

Gen. (lower) Submodular Polyhedron: extreme points

Lemma 2

Given submodular f, (x, c) is an extreme point of $\mathcal{P}_f^{gen}$ if and only if x is an extreme point of $\mathcal{P}_f$ and $c = 0$. Furthermore, for any $y \in \mathbb{R}^n$,

$$\max_{(x,c)\in\mathcal{P}_f^{gen}} [\langle x, y\rangle + c] = \max_{x\in\mathcal{P}_f} \langle x, y\rangle \tag{14}$$

Proof.

Immediately, $\max_{s\in\mathcal{P}_f} w^\top s \leq \max_{(x,c)\in\mathcal{P}_f^{\text{gen}}} [\langle x, w\rangle + c]$. Also, for any $(x, c) \in \mathcal{P}_f^{\text{gen}}$,

$$\max_{s\in\mathcal{P}_f} w^\top s = \sum_i \lambda_i f(S_i^{\sigma_w}) \geq \sum_i \lambda_i [\langle x, 1_{S_i^{\sigma_w}}\rangle + c] \geq \langle x, w\rangle + c, \tag{15}$$

□

Gen. (lower) Submodular Polyhedron: extreme points

Lemma 2

Given submodular f, (x, c) is an extreme point of $\mathcal{P}_f^{gen}$ if and only if x is an extreme point of $\mathcal{P}_f$ and $c = 0$. Furthermore, for any $y \in \mathbb{R}^n$,

$$\max_{(x,c)\in\mathcal{P}_f^{gen}} [\langle x, y\rangle + c] = \max_{x\in\mathcal{P}_f} \langle x, y\rangle \tag{14}$$

Proof.

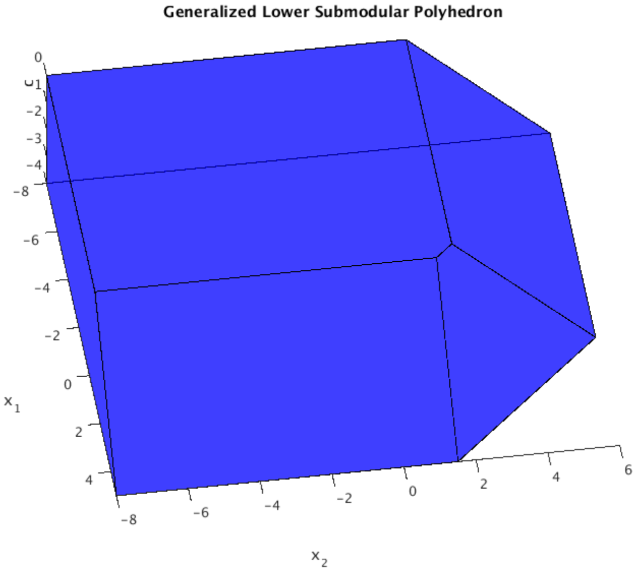
Immediately, $\max_{s\in\mathcal{P}_f} w^\top s \leq \max_{(x,c)\in\mathcal{P}_f^{\text{gen}}} [\langle x, w\rangle + c]$. Also, for any $(x, c) \in \mathcal{P}_f^{\text{gen}}$,

$$\max_{s\in\mathcal{P}_f} w^\top s = \sum_i \lambda_i f(S_i^{\sigma_w}) \geq \sum_i \lambda_i [\langle x, 1_{S_i^{\sigma_w}}\rangle + c] \geq \langle x, w\rangle + c, \tag{15}$$

□

- Also, membership $x \in \mathcal{P}_f^{\text{gen}}$ is still polytime via SFM.

Generalized Polyhedron Visualization



Convex Extensions via the Generalized Polyhedron

- Convex envelope (Vondrák'07, Dughmi'11) of any set function f not nec. submodular, for $w \in [0,1]^n$:

$$\breve{f}(w) = \max_{\phi \in \Phi_f} \phi(w) = \min_{\lambda \in \Lambda_w} \sum_{S \subseteq V} \lambda_S f(S) \tag{16}$$

where

$$\Phi_f = \{\phi : \phi \text{ is convex in } [0,1]^V \text{ and } \phi(1_X) \leq f(X), \forall X \subseteq V\} \tag{17}$$

and (for the r.h.s., a distribution characterization),

$$\Lambda_w = \{\lambda_S, S \subseteq V : \sum_{S \subseteq V} \lambda_S 1_S = w, \sum_{S \subseteq V} \lambda_S = 1, \lambda_S \geq 0\}. \tag{18}$$

Convex Extensions via the Generalized Polyhedron

- Convex envelope (Vondrák'07, Dughmi'11) of any set function f not nec. submodular, for $w \in [0,1]^n$:

$$\breve{f}(w) = \max_{\phi \in \Phi_f} \phi(w) = \min_{\lambda \in \Lambda_w} \sum_{S \subseteq V} \lambda_S f(S) \tag{16}$$

where

$$\Phi_f = \{\phi : \phi \text{ is convex in } [0,1]^V \text{ and } \phi(1_X) \leq f(X), \forall X \subseteq V\} \tag{17}$$

and (for the r.h.s., a distribution characterization),

$$\Lambda_w = \{\lambda_S, S \subseteq V : \sum_{S \subseteq V} \lambda_S 1_S = w, \sum_{S \subseteq V} \lambda_S = 1, \lambda_S \geq 0\}. \tag{18}$$

Lemma 3

Convex extension of f in equation (67) can be expressed as:

$$\breve{f}(w) = \max_{(x,c) \in \mathcal{P}_f^{gen}} [\langle x, w \rangle + c], \forall w \in [0,1]^n \tag{19}$$

Convex Extensions via the Generalized Polyhedron

- Convex envelope (Vondrák'07, Dughmi'11) of any set function f not nec. submodular, for $w \in [0,1]^n$:

$$\breve{f}(w) = \max_{\phi \in \Phi_f} \phi(w) = \min_{\lambda \in \Lambda_w} \sum_{S \subseteq V} \lambda_S f(S) \tag{16}$$

where

$$\Phi_f = \{\phi : \phi \text{ is convex in } [0,1]^V \text{ and } \phi(1_X) \leq f(X), \forall X \subseteq V\} \tag{17}$$

and (for the r.h.s., a distribution characterization),

$$\Lambda_w = \{\lambda_S, S \subseteq V : \sum_{S \subseteq V} \lambda_S 1_S = w, \sum_{S \subseteq V} \lambda_S = 1, \lambda_S \geq 0\}. \tag{18}$$

Proof.

The achieving convex function $\hat{\phi}$ has a tight subgradient (x, d) with $\langle x, y \rangle + d \leq \hat{\phi}(y), \forall y$ and $\langle x, w \rangle + d = \hat{\phi}(w)$. Then $(x, d) \in \mathcal{P}_f^{\text{gen}}$ since $x(X) + d \leq \hat{\phi}(1_X) \leq f(X), \forall X \subseteq V$. □

Convex Extension Submodular Case

- For submodular functions, the convex extension is the Lovász extension and can be expressed:

$$\breve{f}(w) \triangleq \max_{s \in P_f} w^\top s = \max_{(s,c) \in P_f^{\text{gen}}} [w^\top s + c] \tag{19}$$

$$= \sum_{i=1}^{n} w(\sigma_w(i))[f(S_i^{\sigma_w}) - f(S_{i-1}^{\sigma_w}] = \sum_{i=1}^{n} \lambda_i f(S_i^{\sigma_w})$$

with $\sum_i \lambda_i \mathbf{1}_{S_i^{\sigma_w}} = w, \sum_i \lambda_i = 1$.

Convex Extension Submodular Case

- For submodular functions, the convex extension is the Lovász extension and can be expressed:

$$\breve{f}(w) \triangleq \max_{s \in P_f} w^\top s = \max_{(s,c) \in P_f^{\text{gen}}} [w^\top s + c] \tag{19}$$

$$= \sum_{i=1}^{n} w(\sigma_w(i))[f(S_i^{\sigma_w}) - f(S_{i-1}^{\sigma_w}] = \sum_{i=1}^{n} \lambda_i f(S_i^{\sigma_w})$$

with $\sum_i \lambda_i \mathbf{1}_{S_i^{\sigma_w}} = w, \sum_i \lambda_i = 1$.

- For non-submodular functions, these easy expressions do not hold and it is NP-hard to evaluate in general (Vondrák'07, Dughmi'11).

Outline

Other convex aspects of submodular functions

Theorem 3 (Edmonds'70,Lovász'83)

For a submodular function $f : 2^V \to \mathbb{R}$, *we have*

$$\min_{X \subseteq V} f(X) = \min_{x \in [0,1]^n} \breve{f}(x) \tag{20}$$

r.h.s. solution has tight rounding algorithm using simple thresholding.

Lemma 4 (Fujishige'91,'05)

A set $A \subseteq V$ *is a minimizer of* $f : 2^V \to \mathbb{R}$ *if and only if:*

$$\mathbf{0} \in \partial_f(A) \tag{21}$$

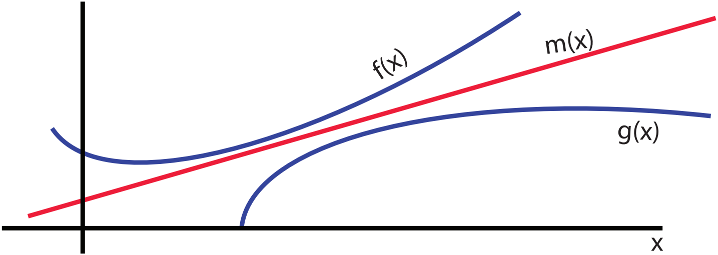
Lemma 5 (Fujishige'91,'05)

A set A *minimizes a submodular function* f *if and only if* $f(A) \leq f(B)$ *for all sets* B *such that* $B \subseteq A$ *or* $A \subseteq B$.

Frank's discrete separation theorem (DST)

Lemma 6 (Frank'82)

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



Fenchel Duality Theorem (FDT)

- Define the Fenchel duals f^* of f, and g^* of g, are respectively convex and concave functions.

$$f^*(x) = \max_{X \subseteq V}[x(X) - f(X)], \quad g^*(x) = \min_{X \subseteq V}[x(X) - g(X)]. \quad (22)$$

Fenchel Duality Theorem (FDT)

- Define the Fenchel duals f^* of f, and g^* of g, are respectively convex and concave functions.

$$f^*(x) = \max_{X \subseteq V}[x(X) - f(X)], \quad g^*(x) = \min_{X \subseteq V}[x(X) - g(X)]. \quad (22)$$

- Then we have:

Fenchel Duality Theorem (FDT)

- Define the Fenchel duals f^* of f, and g^* of g, are respectively convex and concave functions.

$$f^*(x) = \max_{X \subseteq V}[x(X) - f(X)], \quad g^*(x) = \min_{X \subseteq V}[x(X) - g(X)]. \quad (22)$$

- Then we have:

Lemma 7 (Fujishige'05)

Given a submodular function f and a supermodular function g,

$$\min_{X \subseteq V}[f(X) - g(X)] = \max_{x}[g^*(x) - f^*(x)]. \quad (23)$$

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

Outline

1. Introduction
2. Submodularity, Convexity and Concavity
3. Submodularity and Convexity
   - Submodular Polyhedron
   - Convex extension
   - Submodular Subdifferential
   - Generalized lower Submodular Polyhedron
   - Convex aspects of a submodular function
4. Submodularity and Concavity
   - Submodular Upper Polyheron
   - Submodular Superdifferentials
   - Generalized Upper Polyhedron
   - Continuous extensions of a submodular function
   - Concave Aspects of a Submodular Function

Submodularity and Concavity

- We've summarized the convex aspects of a submodular function: Lovász extension, efficient minimization, Franks's DST, subdifferentials and subgradients tight at any point, minimizers of a submodular function form lattice, Fenchel Duality Theorem, all well known.

Submodularity and Concavity

- We've summarized the convex aspects of a submodular function: Lovász extension, efficient minimization, Franks's DST, subdifferentials and subgradients tight at any point, minimizers of a submodular function form lattice, Fenchel Duality Theorem, all well known.
- There are well-known concave aspects of submodular functions as well: The definition $\nabla_j \nabla_k f(X) \leq 0$ where $\nabla_j f(X) = f(j|X)$, concave over modular is submodular, efficient approximate maximization, $1 - 1/e$ or $1/2$ for many problems (Vondrák and many others).

Submodularity and Concavity

- We've summarized the convex aspects of a submodular function: Lovász extension, efficient minimization, Franks's DST, subdifferentials and subgradients tight at any point, minimizers of a submodular function form lattice, Fenchel Duality Theorem, all well known.
- There are well-known concave aspects of submodular functions as well: The definition $\nabla_j \nabla_k f(X) \leq 0$ where $\nabla_j f(X) = f(j|X)$, concave over modular is submodular, efficient approximate maximization, $1 - 1/e$ or $1/2$ for many problems (Vondrák and many others).
- Question: Can one provide a principled theoretical characterization (similar to the the convex aspects of submodular functions), from a concave perspective?

Outline

Submodular Upper Polyhedron

- Define Submodular Upper Polyhedron as follows:

$$\mathcal{P}^f = \{x \in \mathbb{R}^n : x(S) \geq f(S), \forall S \subseteq V\} \tag{24}$$

Submodular Upper Polyhedron

- Define Submodular Upper Polyhedron as follows:

$$\mathcal{P}^f = \{x \in \mathbb{R}^n : x(S) \geq f(S), \forall S \subseteq V\} \tag{24}$$

Lemma 8

Given a submodular function f,

$$\mathcal{P}^f = \{x \in \mathbb{R}^n : x(j) \geq f(j)\} \tag{25}$$

Submodular Upper Polyhedron

- Define Submodular <u>Upper</u> Polyhedron as follows:

$$\mathcal{P}^f = \{x \in \mathbb{R}^n : x(S) \geq f(S), \forall S \subseteq V\} \tag{24}$$

Lemma 8

Given a submodular function f,

$$\mathcal{P}^f = \{x \in \mathbb{R}^n : x(j) \geq f(j)\} \tag{25}$$

Proof.

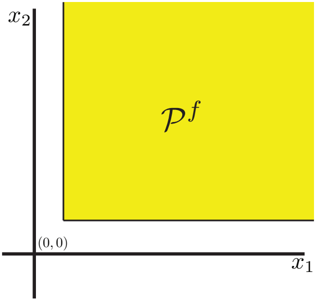
With $x \in \mathcal{P}^f$, for any S, we have

$$x(S) = \sum_{i \in S} x(i) \geq \sum_{i \in S} f(i) \geq f(S) \tag{26}$$

I.e., only the singleton inequalities are irredundant. □

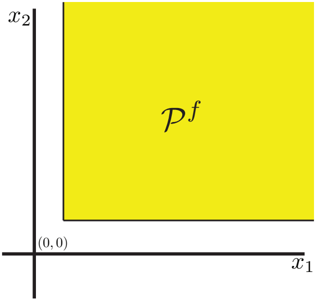
Submodular Upper Polyhedron

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



Submodular Upper Polyhedron

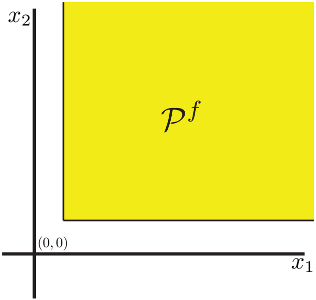
$$\mathcal{P}^f = \{x \in \mathbb{R}^n : x(S) \geq f(S), \forall S \subseteq V\}$$



- Immediate facts:

Submodular Upper Polyhedron

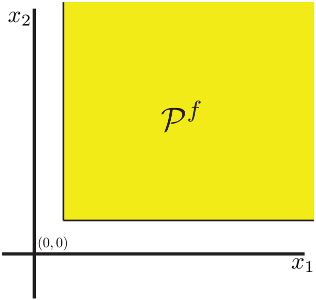
$$\mathcal{P}^f = \{x \in \mathbb{R}^n : x(S) \geq f(S), \forall S \subseteq V\}$$



- Immediate facts:
- Membership problem: $x \in \mathcal{P}^f$ same as $\max_{X \subseteq V} f(X) - x(X) \leq 0$, submodular maximization which is hard,

Submodular Upper Polyhedron

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



- Immediate facts:
- Membership problem: $x \in \mathcal{P}^f$ same as $\max_{X \subseteq V} f(X) - x(X) \leq 0$, submodular maximization which is hard, but in fact same as $\max_{X \subseteq V} \sum_{i \in X} f(i) - x(X) \leq 0$, identical to checking singletons $f(i) - x(i) < 0$.

Outline

Submodular Superdifferentials

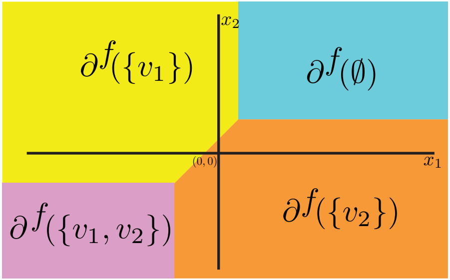
- Analogous to concave functions, we define Submodular Superdifferentials at each $X \subseteq V$:

$$\partial^f(X) = \{x \in \mathbb{R}^n : f(Y) - x(Y) \leq f(X) - x(X), \forall Y \subseteq V\} \quad (27)$$

Submodular Superdifferentials

- Analogous to concave functions, we define Submodular Superdifferentials at each $X \subseteq V$:

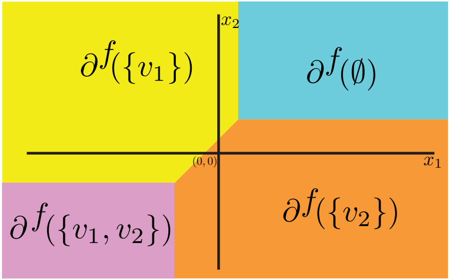
$$\partial^f(X) = \{x \in \mathbb{R}^n : f(Y) - x(Y) \leq f(X) - x(X), \forall Y \subseteq V\} \quad (27)$$



Submodular Superdifferentials

- Analogous to concave functions, we define Submodular Superdifferentials at each $X \subseteq V$:

$$\partial^f(X) = \{x \in \mathbb{R}^n : f(Y) - x(Y) \leq f(X) - x(X), \forall Y \subseteq V\} \quad (27)$$



- Each $g_X \in \partial^f(X)$ defines modular upper bound of f tight at X:

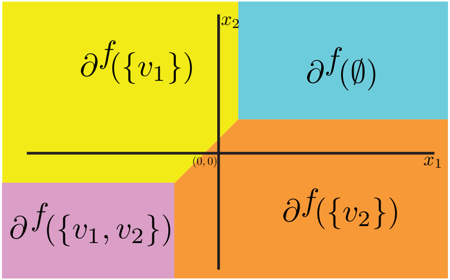
$$\partial_f(X) = \{x \in \mathbb{R}^n : f(Y) \leq f(X) - x(X) + x(Y); \forall Y \subseteq V\}$$

so $m^X(Y) \triangleq f(X) - g_X(X) + g_X(Y) \geq f(Y)$ and $m_X(X) = f(X)$.

Submodular Superdifferentials

- Analogous to concave functions, we define Submodular Superdifferentials at each $X \subseteq V$:

$$\partial^f(X) = \{x \in \mathbb{R}^n : f(Y) - x(Y) \leq f(X) - x(X), \forall Y \subseteq V\} \quad (27)$$



- Each $g_X \in \partial^f(X)$ defines modular upper bound of f tight at X:

$$\partial_f(X) = \{x \in \mathbb{R}^n : f(Y) \leq f(X) - x(X) + x(Y); \forall Y \subseteq V\}$$

so $m^X(Y) \triangleq f(X) - g_X(X) + g_X(Y) \geq f(Y)$ and $m_X(X) = f(X)$.

- m_X is (typically not a normalized) modular function.

Submodular Superdifferential Redundancy

- Define three polyhedra:

$$\partial_1^f(X) = \{x \in \mathbb{R}^n : f(Y) - x(Y) \leq f(X) - x(X), \forall Y \subseteq X\} \quad (28)$$

$$\partial_2^f(X) = \{x \in \mathbb{R}^n : f(Y) - x(Y) \leq f(X) - x(X), \forall Y \supseteq X\} \quad (29)$$

$$\partial_3^f(X) = \{x \in \mathbb{R}^n : f(Y) - x(Y) \leq f(X) - x(X), \forall Y : Y \not\subseteq X, Y \not\supseteq X\}$$

Submodular Superdifferential Redundancy

- Define three polyhedra:

$$\partial_1^f(X) = \{x \in \mathbb{R}^n : f(Y) - x(Y) \leq f(X) - x(X), \forall Y \subseteq X\} \tag{28}$$

$$\partial_2^f(X) = \{x \in \mathbb{R}^n : f(Y) - x(Y) \leq f(X) - x(X), \forall Y \supseteq X\} \tag{29}$$

$$\partial_3^f(X) = \{x \in \mathbb{R}^n : f(Y) - x(Y) \leq f(X) - x(X), \forall Y : Y \not\subseteq X, Y \not\supseteq X\}$$

- Immediately $\partial^f(X) = \partial_1^f(X) \cap \partial_2^f(X) \cap \partial_3^f(X)$. Also, we have

Submodular Superdifferential Redundancy

- Define three polyhedra:

$$\partial_1^f(X) = \{x \in \mathbb{R}^n : f(Y) - x(Y) \le f(X) - x(X), \forall Y \subseteq X\} \quad (28)$$

$$\partial_2^f(X) = \{x \in \mathbb{R}^n : f(Y) - x(Y) \le f(X) - x(X), \forall Y \supseteq X\} \quad (29)$$

$$\partial_3^f(X) = \{x \in \mathbb{R}^n : f(Y) - x(Y) \le f(X) - x(X), \forall Y : Y \not\subseteq X, Y \not\supseteq X\}$$

- Immediately $\partial^f(X) = \partial_1^f(X) \cap \partial_2^f(X) \cap \partial_3^f(X)$. Also, we have

Lemma 9

For submodular f, $\partial_1^f(X)$ and $\partial_2^f(X)$'s irredundant representation is:

$$\partial_1^f(X) = \{x \in \mathbb{R}^n : f(j|X \setminus j) \ge x(j), \forall j \in X\} \quad (30)$$

$$\partial_2^f(X) = \{x \in \mathbb{R}^n : f(j|X) \le x(j), \forall j \notin X\}. \quad (31)$$

Submodular Superdifferential Redundancy

- Define three polyhedra:

$$\partial_1^f(X) = \{x \in \mathbb{R}^n : f(Y) - x(Y) \le f(X) - x(X), \forall Y \subseteq X\} \quad (28)$$

$$\partial_2^f(X) = \{x \in \mathbb{R}^n : f(Y) - x(Y) \le f(X) - x(X), \forall Y \supseteq X\} \quad (29)$$

$$\partial_3^f(X) = \{x \in \mathbb{R}^n : f(Y) - x(Y) \le f(X) - x(X), \forall Y : Y \not\subseteq X, Y \not\supseteq X\}$$

- Immediately $\partial^f(X) = \partial_1^f(X) \cap \partial_2^f(X) \cap \partial_3^f(X)$. Also, we have

Lemma 9

For submodular f, $\partial_1^f(X)$ and $\partial_2^f(X)$'s irredundant representation is:

$$\partial_1^f(X) = \{x \in \mathbb{R}^n : f(j|X \setminus j) \ge x(j), \forall j \in X\} \quad (30)$$

$$\partial_2^f(X) = \{x \in \mathbb{R}^n : f(j|X) \le x(j), \forall j \notin X\}. \quad (31)$$

Proof.

Eq. (28) $\Leftrightarrow$ Eq. (30). Assuming only $f(j|X \setminus j) \ge x(j)$, gives $x(X) - x(Y) = x(X \setminus Y) = \sum_{j \in X \setminus Y} x(j) \le \sum_{j \in X \setminus Y} f(j|X \setminus j) \le f(X|Y)$ when $X \supseteq Y$. Eq. (29) $\Leftrightarrow$ Eq. (31) similar. □

Submodular Superdifferential 2D extreme points

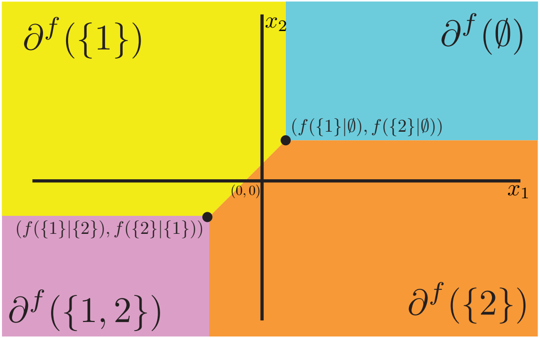
- In 2D, superdifferential at v_1 takes the form:

$$\partial^f(v_1) = \{x \in \mathbb{R}^2 : x_1 \leq f(v_1), \quad (32)$$

$$x_2 \geq f(v_2|v_1), \quad (33)$$

$$x_1 - x_2 \leq f(v_1) - f(v_2)\} \quad (34)$$

so extreme points are at indicated below:



Submodular Superdifferential 2D extreme points

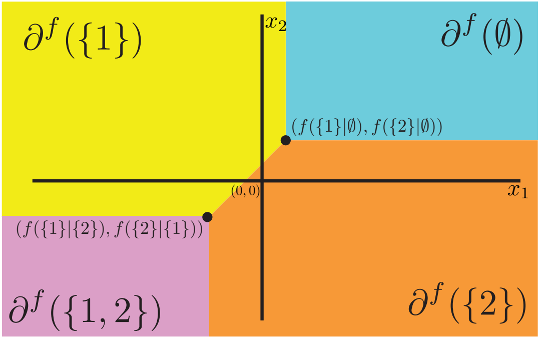
- In 2D, superdifferential at v_1 takes the form:

$$\partial^f(v_1) = \{x \in \mathbb{R}^2 : x_1 \le f(v_1), \tag{32}$$

$$x_2 \ge f(v_2|v_1), \tag{33}$$

$$x_1 - x_2 \le f(v_1) - f(v_2)\} \tag{34}$$

so extreme points are at indicated below:



- Can analytically characterize superdifferential structure in 2D.

Submodular Superdifferential 3D extreme points

- 3D (with $V = \{a, b, c\}$, superdifferential at a irredundantly expressed:

$$\partial^f(a) = \{x \in \mathbb{R}^3 : x(a) \leq f(a), (Y = \emptyset) \tag{35}$$

$$x(b) \geq f(b|a), (Y = \{a, b\}) \tag{36}$$

$$x(b) - x(a) \geq f(b) - f(a), (Y = \{b\}) \tag{37}$$

$$x(c) \geq f(c|a), (Y = \{a, c\}) \tag{38}$$

$$x(c) - x(a) \geq f(c) - f(a), (Y = \{c\}) \tag{39}$$

$$x(b) + x(c) - x(a) \geq f(b, c) - f(a), (Y = \{b, c\})\} \tag{40}$$

Submodular Superdifferential 3D extreme points

- 3D (with $V = \{a, b, c\}$, superdifferential at a irredundantly expressed:

$$\partial^f(a) = \{x \in \mathbb{R}^3 : x(a) \leq f(a), (Y = \emptyset) \tag{35}$$
$$x(b) \geq f(b|a), (Y = \{a, b\}) \tag{36}$$
$$x(b) - x(a) \geq f(b) - f(a), (Y = \{b\}) \tag{37}$$
$$x(c) \geq f(c|a), (Y = \{a, c\}) \tag{38}$$
$$x(c) - x(a) \geq f(c) - f(a), (Y = \{c\}) \tag{39}$$
$$x(b) + x(c) - x(a) \geq f(b, c) - f(a), (Y = \{b, c\})\} \tag{40}$$

- Immediately, $(f(a), f(b), f(c))$ is an extreme point.

Submodular Superdifferential 3D extreme points

- 3D (with $V = \{a, b, c\}$, superdifferential at a irredundantly expressed:

$$\begin{aligned}
\partial^f(a) = \{x \in \mathbb{R}^3 : & x(a) \leq f(a), (Y = \emptyset) && (35)\\
& x(b) \geq f(b|a), (Y = \{a, b\}) && (36)\\
& x(b) - x(a) \geq f(b) - f(a), (Y = \{b\}) && (37)\\
& x(c) \geq f(c|a), (Y = \{a, c\}) && (38)\\
& x(c) - x(a) \geq f(c) - f(a), (Y = \{c\}) && (39)\\
& x(b) + x(c) - x(a) \geq f(b, c) - f(a), (Y = \{b, c\})\} && (40)
\end{aligned}$$

- Immediately, $(f(a), f(b), f(c))$ is an extreme point.
- Other extreme points not possible to identify without knowing more about the function (e.g., if $f(b|c) < f(b|a)$ or if $f(b|c) > f(b|a)$).

Submodular Superdifferential 3D extreme points

- 3D (with $V = \{a, b, c\}$, superdifferential at a irredundantly expressed:

$$\begin{aligned}
\partial^f(a) = \{x \in \mathbb{R}^3 : & x(a) \le f(a), (Y = \emptyset) && (35)\\
& x(b) \ge f(b|a), (Y = \{a, b\}) && (36)\\
& x(b) - x(a) \ge f(b) - f(a), (Y = \{b\}) && (37)\\
& x(c) \ge f(c|a), (Y = \{a, c\}) && (38)\\
& x(c) - x(a) \ge f(c) - f(a), (Y = \{c\}) && (39)\\
& x(b) + x(c) - x(a) \ge f(b, c) - f(a), (Y = \{b, c\})\} && (40)
\end{aligned}$$

- Immediately, $(f(a), f(b), f(c))$ is an extreme point.
- Other extreme points not possible to identify without knowing more about the function (e.g., if $f(b|c) < f(b|a)$ or if $f(b|c) > f(b|a)$).
- Superdifferential much harder to characterize than subdifferential.

Superdifferential membership problem is hard

Lemma 10

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

Proof.

$y \in \partial^f(Y)$ same as asking $\max_{X \subseteq V}[f(X) - y(X)] \leq f(Y) - y(Y)$, i.e., is Y is a maximizer of $f(X) - y(X)$ for a given vector y? Decision version of submodular maximization, NP hard when $\emptyset \subset Y \subset V$. □

Superdifferential membership problem is hard

Lemma 10

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

Proof.

$y \in \partial^f(Y)$ same as asking $\max_{X \subseteq V}[f(X) - y(X)] \leq f(Y) - y(Y)$, i.e., is Y is a maximizer of $f(X) - y(X)$ for a given vector y? Decision version of submodular maximization, NP hard when $\emptyset \subset Y \subset V$. □

- Hence, since membership problem is hard, linear program over it is hard (Grotschel, Lovász, Schrijver,'84).

Superdifferential membership problem is hard

Lemma 10

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

Proof.

$y \in \partial^f(Y)$ same as asking $\max_{X \subseteq V}[f(X) - y(X)] \leq f(Y) - y(Y)$, i.e., is Y is a maximizer of $f(X) - y(X)$ for a given vector y? Decision version of submodular maximization, NP hard when $\emptyset \subset Y \subset V$. □

- Hence, since membership problem is hard, linear program over it is hard (Grotschel, Lovász, Schrijver,'84).
- Empty set and ground set easy, however, and are characterized as:

$$\partial^f(\emptyset) = \{x \in \mathbb{R}^n : f(j) \leq x(j), \forall j \in V\} \tag{41}$$

$$\partial^f(V) = \{x \in \mathbb{R}^n : f(j|V \setminus j) \geq x(j), \forall j \in V\} \tag{42}$$

with $\partial^f(\emptyset) = \mathcal{P}^f$ and $\partial^f(V) = \mathcal{P}^{f^\#}$ where $f^\#(X) = f(V) - f(V \setminus X)$ is the submodular dual of f.

Superdifferential outer bounds

- Recall from Lemma 9 that $\partial^f_1(X) = \{x \in \mathbb{R}^n : f(j|X\backslash j) \geq x(j), \forall j \in X\}$ and $\partial^f_2(X) = \{x \in \mathbb{R}^n : f(j|X) \leq x(j), \forall j \notin X\}$ are easily characterized, and
$\partial^f_3(X) = \{x \in \mathbb{R}^n : f(Y) - x(Y) \leq f(X) - x(X), \} \forall Y : Y \not\subseteq X, Y \not\supseteq X\}$
is what makes $\partial^f(X)$ hard.

Superdifferential outer bounds

- Recall from Lemma 9 that $\partial^f_1(X) = \{x \in \mathbb{R}^n : f(j|X\setminus j) \geq x(j), \forall j \in X\}$ and $\partial^f_2(X) = \{x \in \mathbb{R}^n : f(j|X) \leq x(j), \forall j \notin X\}$ are easily characterized, and
$\partial^f_3(X) = \{x \in \mathbb{R}^n : f(Y) - x(Y) \leq f(X) - x(X), \}\forall Y : Y \not\subseteq X, Y \not\supseteq X\}$
is what makes $\partial^f(X)$ hard.
- Define outer bound $\partial^f_{3,\Delta(k,l)}(X) \supseteq \partial^f_3(X)$:

$$\partial^f_{3,\Delta(k,l)}(X) = \{x \in \mathbb{R}^n : f(Y) - x(Y) \leq f(X) - x(X), \forall Y : Y \not\subseteq X, Y \not\supseteq X, |Y\setminus X| \leq k-1, |X\setminus Y| \leq l-1\} \tag{43}$$

and then

$$\partial^f_{\Delta(k,l)}(X) = \partial^f_1(X) \cap \partial^f_2(X) \cap \partial^f_{3,\Delta(k,l)}(X). \tag{44}$$

Superdifferential outer bound containment

Theorem 11

For a submodular function f:

1. $\partial^f_{\Delta(1,1)}(X) = \partial^f_1(X) \cap \partial^f_2(X)$
2. $\forall 1 \leq k' \leq k, 1 \leq l' \leq l, \partial^f(X) \subseteq \partial^f_{\Delta(k,l)}(X) \subseteq \partial^f_{\Delta(k',l')}(X) \subseteq \partial^f_{\Delta(1,1)}(X)$
3. $\partial^f_{\Delta(n,n)}(X) = \partial^f(X)$.

Superdifferential outer bound containment

Theorem 11

For a submodular function f:

1. $\partial^f_{\Delta(1,1)}(X) = \partial^f_1(X) \cap \partial^f_2(X)$
2. $\forall 1 \leq k' \leq k, 1 \leq l' \leq l, \partial^f(X) \subseteq \partial^f_{\Delta(k,l)}(X) \subseteq \partial^f_{\Delta(k',l')}(X) \subseteq \partial^f_{\Delta(1,1)}(X)$
3. $\partial^f_{\Delta(n,n)}(X) = \partial^f(X)$.

- We call $\partial^f_{\Delta(1,1)}(X) = \{x \in \mathbb{R}^n : x(j) \leq f(j|X\setminus j) \forall j \in X, x(j) \geq f(j|X) \forall j \notin X\}$ the local superdifferential approximation.

Visualization of the Superdifferential its outer bounds

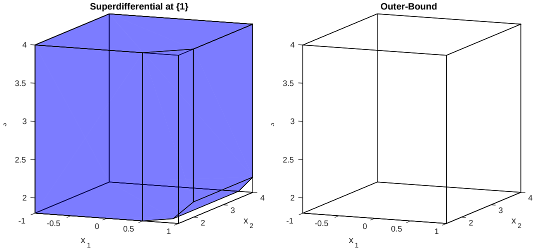


Figure: A visualization of the outer bounds of the superdifferential. The first figure (top left) is the submodular supergradient $\partial^f(X)$, while the second one (top) is the outer bound $\partial^f_{\Delta(1,1)}(X)$.

Achievable Submodular Tight Supergradients

- Nemhauser, Wolsey, & Fisher'78 characterized submodularity with either of the following, $\forall X, Y$:

$$f(Y) \leq f(X) - \sum_{j \in X \setminus Y} f(j|X \setminus j) + \sum_{j \in Y \setminus X} f(j|X \cap Y), \tag{45}$$

$$f(Y) \leq f(X) - \sum_{j \in X \setminus Y} f(j|X \cup Y \setminus j) + \sum_{j \in Y \setminus X} f(j|X) \tag{46}$$

Achievable Submodular Tight Supergradients

- Nemhauser, Wolsey, & Fisher'78 characterized submodularity with either of the following, $\forall X, Y$:

$$f(Y) \leq f(X) - \sum_{j \in X \setminus Y} f(j|X \setminus j) + \sum_{j \in Y \setminus X} f(j|X \cap Y), \tag{45}$$

$$f(Y) \leq f(X) - \sum_{j \in X \setminus Y} f(j|X \cup Y \setminus j) + \sum_{j \in Y \setminus X} f(j|X) \tag{46}$$

- Using submodularity, these can be further loosened as follows (Ahmed & Atamtürk,'09; Jegelka & B.'09; Iyer, Jegelka, B.'12,'13):

$$f(Y) \leq f(X) - \sum_{j \in X \setminus Y} f(j|X - \{j\}) + \sum_{j \in Y \setminus X} f(j|\emptyset) \tag{47}$$

$$f(Y) \leq f(X) - \sum_{j \in X \setminus Y} f(j|V - \{j\}) + \sum_{j \in Y \setminus X} f(j|X) \tag{48}$$

$$f(Y) \leq f(X) - \sum_{j \in X \setminus Y} f(j|V - \{j\}) + \sum_{j \in Y \setminus X} f(j|\emptyset). \tag{49}$$

Achievable Submodular Tight Supergradients

- Each of the bounds above offers a supergradient.

Achievable Submodular Tight Supergradients

- Each of the bounds above offers a supergradient.
- For example, starting from Equation (47),

$$f(Y) \leq f(X) - \sum_{j \in X \setminus Y} f(j|X \setminus \{j\}) + \sum_{j \in Y \setminus X} f(j|\emptyset) \tag{50}$$

$$= f(X) - \sum_{j \in X} f(j|X \setminus \{j\}) + \sum_{j \in X \cap Y} f(j|X \setminus \{j\}) + \sum_{j \in Y \setminus X} f(j|\emptyset) \tag{51}$$

$$= f(X) - \hat{g}_X(X) + \hat{g}_X(Y) \tag{52}$$

where $\hat{g}_X$ is defined as:

$$\hat{g}_X(j) = \begin{cases} f(j|X - j) & \text{if } j \in X \\ f(j) & \text{if } j \notin X \end{cases} \tag{53}$$

Achievable Submodular Tight Supergradients

- Each of the bounds above offers a supergradient.
- For example, starting from Equation (47),

$$f(Y) \leq f(X) - \sum_{j \in X \setminus Y} f(j|X \setminus \{j\}) + \sum_{j \in Y \setminus X} f(j|\emptyset) \quad (50)$$

$$= f(X) - \sum_{j \in X} f(j|X \setminus \{j\}) + \sum_{j \in X \cap Y} f(j|X \setminus \{j\}) + \sum_{j \in Y \setminus X} f(j|\emptyset) \quad (51)$$

$$= f(X) - \hat{g}_X(X) + \hat{g}_X(Y) \quad (52)$$

where $\hat{g}_X$ is defined as:

$$\hat{g}_X(j) = \begin{cases} f(j|X - j) & \text{if } j \in X \\ f(j) & \text{if } j \notin X \end{cases} \quad (53)$$

- Thus, $\hat{g}_X \in \partial^f(X)$, proving the non-emptiness of $\partial^f(X)$ for any $X \subseteq V$.

The Three Supergradients (Iyer et al, 2013)

Define three vectors $\in \mathbb{R}^V$ as follows:

$$\hat{g}_X(j) = \begin{cases} f(j|X - j) & \text{if } j \in X \\ f(j) & \text{if } j \notin X \end{cases} \tag{54}$$

$$\check{g}_X(j) = \begin{cases} f(j|V - j) & \text{if } j \in X \\ f(j|X) & \text{if } j \notin X \end{cases} \tag{55}$$

$$\bar{g}_X(j) = \begin{cases} f(j|V - j) & \text{if } j \in X \\ f(j) & \text{if } j \notin X \end{cases} \tag{56}$$

Theorem 12

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

Visualization of the Three Supergradients

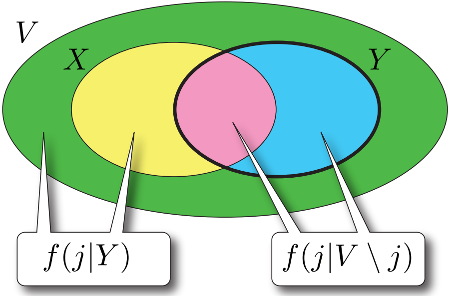
- We call these supergradients grow, shrink, and bar

Visualization of the Three Supergradients

- We call these supergradients grow, shrink, and bar

Grow:

$$\hat{g}_Y(j) = \begin{cases} f(j|Y) & \text{for } j \notin Y \\ f(j|V\backslash\{j\}) & \text{for } j \in Y \end{cases}$$

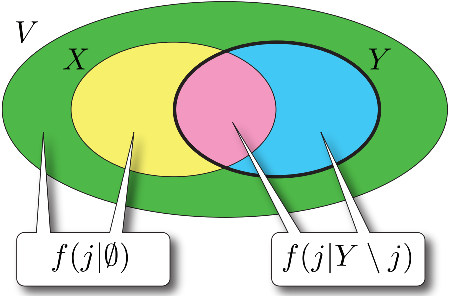


Visualization of the Three Supergradients

- We call these supergradients grow, shrink, and bar

Shrink:

$$\breve{g}_Y(j) = \begin{cases} f(j|\emptyset) & \text{for } j \notin Y \\ f(j|Y\setminus\{j\}) & \text{for } j \in Y \end{cases}$$

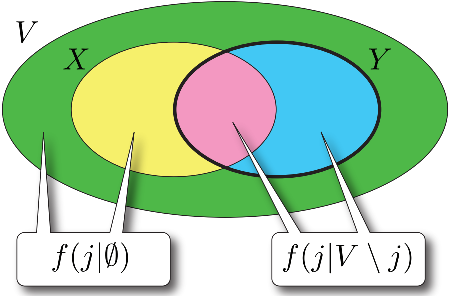


Visualization of the Three Supergradients

- We call these supergradients grow, shrink, and bar

Bar:

$$\bar{g}_Y(j) = \begin{cases} f(j|\emptyset) & \text{for } j \notin Y \\ f(j|V \setminus \{j\}) & \text{for } j \in Y \end{cases}$$

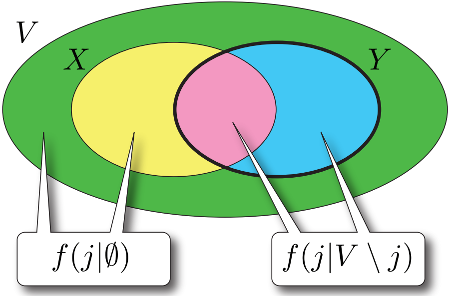


Visualization of the Three Supergradients

- We call these supergradients grow, shrink, and bar

Bar:

$$\bar{g}_Y(j) = \begin{cases} f(j|\emptyset) & \text{for } j \notin Y \\ f(j|V\setminus\{j\}) & \text{for } j \in Y \end{cases}$$



- Modular upper bound: $m^{g_Y}(X) = f(Y) + g_Y(X) - g_Y(Y) \leq f(X)$.

Single Extreme Superdifferential Inner Bounds

- First, define the following two polyhedra:

$$\partial^f_\emptyset(X) = \{x \in \mathbb{R}^n : f(j) \leq x(j), \forall j \notin X\}, \quad (57)$$

$$\partial^f_V(X) = \{x \in \mathbb{R}^n : f(j|V\backslash j) \geq x(j), \forall j \in X\}. \quad (58)$$

Single Extreme Superdifferential Inner Bounds

- First, define the following two polyhedra:

$$\partial^f_\emptyset(X) = \{x \in \mathbb{R}^n : f(j) \leq x(j), \forall j \notin X\}, \quad (57)$$
$$\partial^f_V(X) = \{x \in \mathbb{R}^n : f(j|V \setminus j) \geq x(j), \forall j \in X\}. \quad (58)$$

- Define:

$$\partial^f_{i,1}(X) = \partial^f_1(X) \cap \partial^f_V(X) \quad (59)$$
$$= \{x \in \mathbb{R}^n : f(j|X \setminus j) \geq x(j), \forall j \in X \text{ and } f(j) \leq x(j), \forall j \notin X\}$$
$$\partial^f_{i,2}(Y) = \partial^f_2(Y) \cap \partial^f_\emptyset(Y) \quad (60)$$
$$= \{x \in \mathbb{R}^n : f(j|V \setminus j) \geq x(j), \forall j \in X \text{ and } f(j|X) \leq x(j), \forall j \notin X\}$$
$$\partial^f_{i,3}(Y) = \partial^f_V(Y) \cap \partial^f_\emptyset(Y) \quad (61)$$
$$= \{x \in \mathbb{R}^n : f(j|V \setminus j) \geq x(j), \forall j \in X \text{ and } f(j) \leq x(j), \forall j \notin X\}.$$

Single Extreme Superdifferential Inner Bounds

- First, define the following two polyhedra:

$$\partial^f_\emptyset(X) = \{x \in \mathbb{R}^n : f(j) \le x(j), \forall j \notin X\}, \quad (57)$$

$$\partial^f_V(X) = \{x \in \mathbb{R}^n : f(j|V \setminus j) \ge x(j), \forall j \in X\}. \quad (58)$$

- Define:

$$\partial^f_{i,1}(X) = \partial^f_1(X) \cap \partial^f_V(X) \quad (59)$$
$$= \{x \in \mathbb{R}^n : f(j|X \setminus j) \ge x(j), \forall j \in X \text{ and } f(j) \le x(j), \forall j \notin X\}$$

$$\partial^f_{i,2}(Y) = \partial^f_2(Y) \cap \partial^f_\emptyset(Y) \quad (60)$$
$$= \{x \in \mathbb{R}^n : f(j|V \setminus j) \ge x(j), \forall j \in X \text{ and } f(j|X) \le x(j), \forall j \notin X\}$$

$$\partial^f_{i,3}(Y) = \partial^f_V(Y) \cap \partial^f_\emptyset(Y) \quad (61)$$
$$= \{x \in \mathbb{R}^n : f(j|V \setminus j) \ge x(j), \forall j \in X \text{ and } f(j) \le x(j), \forall j \notin X\}.$$

- We have: $\hat{g}_Y$ is an extreme point of $\partial^f_{i,1}(Y)$, $\check{g}_Y$ and extreme point of $\partial^f_{i,2}(Y)$, and $\bar{g}_Y$ an extreme point of $\partial^f_{i,3}(Y)$.

Single Extreme Superdifferential Inner Bounds

- First, define the following two polyhedra:

$$\partial^f_\emptyset(X) = \{x \in \mathbb{R}^n : f(j) \leq x(j), \forall j \notin X\}, \quad (57)$$
$$\partial^f_V(X) = \{x \in \mathbb{R}^n : f(j|V\backslash j) \geq x(j), \forall j \in X\}. \quad (58)$$

- Define:

$$\partial^f_{i,1}(X) = \partial^f_1(X) \cap \partial^f_V(X) \quad (59)$$
$$= \{x \in \mathbb{R}^n : f(j|X\backslash j) \geq x(j), \forall j \in X \text{ and } f(j) \leq x(j), \forall j \notin X\}$$
$$\partial^f_{i,2}(Y) = \partial^f_2(Y) \cap \partial^f_\emptyset(Y) \quad (60)$$
$$= \{x \in \mathbb{R}^n : f(j|V\backslash j) \geq x(j), \forall j \in X \text{ and } f(j|X) \leq x(j), \forall j \notin X\}$$
$$\partial^f_{i,3}(Y) = \partial^f_V(Y) \cap \partial^f_\emptyset(Y) \quad (61)$$
$$= \{x \in \mathbb{R}^n : f(j|V\backslash j) \geq x(j), \forall j \in X \text{ and } f(j) \leq x(j), \forall j \notin X\}.$$

- We have: $\hat{g}_Y$ is an extreme point of $\partial^f_{i,1}(Y)$, $\check{g}_Y$ and extreme point of $\partial^f_{i,2}(Y)$, and $\bar{g}_Y$ an extreme point of $\partial^f_{i,3}(Y)$.
- All are simple polyhedra, with a single extreme point.

Superdifferential Inner Bounds

- Define a combination of polyhedra as follows:

$$\partial^f_{i,(1,2)}(Y) = \text{conv}(\partial^f_{i,1}(Y), \partial^f_{i,2}(Y)) \tag{62}$$

where conv(., .) is the convex combination of two polyhedra.

Superdifferential Inner Bounds

- Define a combination of polyhedra as follows:

$$\partial^f_{i,(1,2)}(Y) = \text{conv}(\partial^f_{i,1}(Y), \partial^f_{i,2}(Y)) \tag{62}$$

  where conv(., .) is the convex combination of two polyhedra.
- Then we have a simple DAG of inclusory properties:

Superdifferential Inner Bounds

- Define a combination of polyhedra as follows:

$$\partial^f_{i,(1,2)}(Y) = \text{conv}(\partial^f_{i,1}(Y), \partial^f_{i,2}(Y)) \tag{62}$$

where conv(., .) is the convex combination of two polyhedra.

- Then we have a simple DAG of inclusory properties:

Lemma 13

Given a submodular function f,

$$\partial^f_{i,3}(Y) \begin{array}{c} \subsetneq \partial^f_{i,2}(Y) \subsetneq \\ \subsetneq \partial^f_{i,1}(Y) \subsetneq \end{array} \partial^f_{i,(1,2)}(Y) \subseteq \partial^f(Y) \tag{63}$$

Superdifferentials and $M^\natural$-concave functions

Lemma 14

Given a submodular function f which is $M^\natural$-concave (Murota,'96,'97,'03) on $\{0,1\}^V$, its superdifferential satisfies,

$$\partial^f(X) = \partial^f_{\Delta(2,2)}(X) \tag{64}$$

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

Proof.

Theorem 6.61 in Murota'03,'10 for an $M^\natural$ convex function (which is supermodular), its subdifferential can be expressed considering only sets Y satisfying $|X \setminus Y| \leq 1, |Y \setminus X| \leq 1$ (Hamming distance less than 2). Superdifferential of a $M^\natural$ concave function (which is supermodular) can be expressed with the same number of inequalities, and the corresponding polyhedron is $\partial^f_{\Delta(2,2)}(X)$. □

Visualization of the Superdifferential its inner and outer bounds

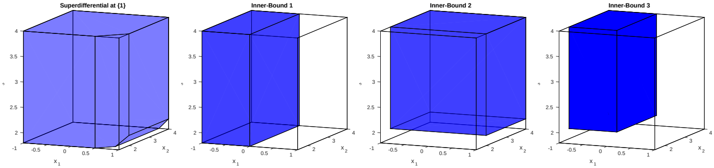


Figure: A visualization of the inner and outer bounds of the superdifferential. The first figure is the submodular supergradient $\partial^f(X)$, while the next three figures show the inner bounds $\partial^f_{i,1}(X), \partial^f_{i,2}(X)$ and $\partial^f_{i,3}(X)$ (marked as inner bounds 1, 2 and 3 respectively).

Outline

Generalized Upper Polyhedron

- Define a generalization of the submodular upper polyhedron $\mathcal{P}^f_{\text{gen}} \subseteq \mathbb{R}^{|V|+1}$:

$$\mathcal{P}^f_{\text{gen}} = \{(x, c), x \in \mathbb{R}^n, c \in \mathbb{R} : x(X) + c \geq f(X), \forall X \subseteq V\} \quad (65)$$

Generalized Upper Polyhedron

- Define a generalization of the submodular upper polyhedron $\mathcal{P}^f_{\text{gen}} \subseteq \mathbb{R}^{|V|+1}$:

$$\mathcal{P}^f_{\text{gen}} = \{(x, c), x \in \mathbb{R}^n, c \in \mathbb{R} : x(X) + c \geq f(X), \forall X \subseteq V\} \quad (65)$$

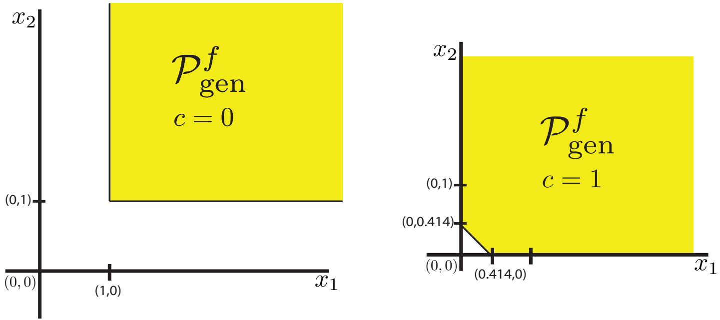
- For normalized submodular functions ($f(\emptyset) = 0$), we have $c \geq 0$.

Generalized Upper Polyhedron

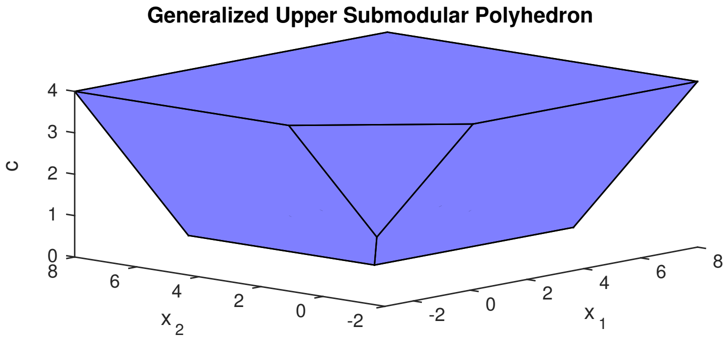
- Define a generalization of the submodular upper polyhedron $\mathcal{P}^f_{\text{gen}} \subseteq \mathbb{R}^{|V|+1}$:

$$\mathcal{P}^f_{\text{gen}} = \{(x, c), x \in \mathbb{R}^n, c \in \mathbb{R} : x(X) + c \geq f(X), \forall X \subseteq V\} \quad (65)$$

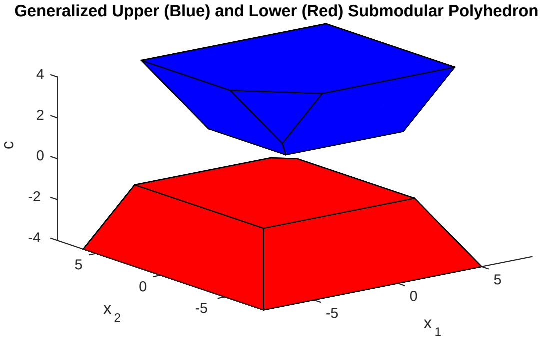
- For normalized submodular functions ($f(\emptyset) = 0$), we have $c \geq 0$.
- Example: $|V| = 2$, $f(X) = \sqrt{|X|}$, two slices $c = 0$ and $c = 1$.



Visualization of the Generalized Upper Submodular Polyhedron



Generalized Upper v.s Lower Submodular Polyhedron



Generalized Upper Polyhedron: LP and Membership

Lemma 15

For submodular function f, and a $y \in \mathbb{R}^n$,

$$\min_{(x,c)\in\mathcal{P}^f_{gen}} \langle x, y\rangle + c = \min\{ \min_{x\in\partial^f(X)} \langle x, y\rangle + f(X) - x(X) \mid X \subseteq V\}. \tag{66}$$

So characterizing generalized submodular upper polyhedron is cast in terms of characterizing the superdifferential.

Generalized Upper Polyhedron: LP and Membership

Lemma 15

For submodular function f, and a $y \in \mathbb{R}^n$,

$$\min_{(x,c)\in\mathcal{P}^f_{gen}} \langle x, y\rangle + c = \min\{\min_{x\in\partial^f(X)} \langle x, y\rangle + f(X) - x(X) \mid X \subseteq V\}. \quad (66)$$

So characterizing generalized submodular upper polyhedron is cast in terms of characterizing the superdifferential. Moreover,

Lemma 16

The generalized submodular upper polyhedron membership problem for submodular f, i.e., is $(x, c) \in \mathcal{P}^f_{gen}$, is NP hard for $c > 0$. Furthermore, the LP $\min_{(x,c)\in\mathcal{P}^f_{gen}} \langle x, y\rangle + c$ is also NP hard.

Proof.

Reduce to submodular max, and use Grotschel, Lovász, Schrijver,'84. □

Outline

Concave Extension

- Concave extension (Vondrák'07, Dughmi'11) of any set function f not nec. submodular, for $w \in [0,1]^n$:

$$\breve{f}(w) = \min_{\phi \in \Phi_f} \phi(w) \tag{67}$$

where

$$\Psi_f = \{\psi : \psi \text{ is concave in } [0,1]^V \text{ and } \psi(1_X) \geq f(X), \forall X \subseteq V\} \tag{68}$$

Concave Extension

- Concave extension (Vondrák'07, Dughmi'11) of any set function f not nec. submodular, for $w \in [0,1]^n$:

$$\breve{f}(w) = \min_{\phi \in \Phi_f} \phi(w) \tag{67}$$

  where

$$\Psi_f = \{\psi : \psi \text{ is concave in } [0,1]^V \text{ and } \psi(1_X) \geq f(X), \forall X \subseteq V\} \tag{68}$$

- We can express this using the generalized submodular upper polyhedron:

Concave Extension

- Concave extension (Vondrák'07, Dughmi'11) of any set function f not nec. submodular, for $w \in [0,1]^n$:

$$\breve{f}(w) = \min_{\phi \in \Phi_f} \phi(w) \quad (67)$$

where

$$\Psi_f = \{\psi : \psi \text{ is concave in } [0,1]^V \text{ and } \psi(1_X) \geq f(X), \forall X \subseteq V\} \quad (68)$$

- We can express this using the generalized submodular upper polyhedron:

Lemma 17

The concave extension of any set function f can be expressed as:

$$\hat{f}(w) = \min_{(y,c) \in \mathcal{P}^f_{gen}} \langle y, w \rangle + c, \forall w \in [0,1]^{|V|} \quad (69)$$

Concave Extension

- Concave extension (Vondrák'07, Dughmi'11) of any set function f not nec. submodular, for $w \in [0,1]^n$:

$$\breve{f}(w) = \min_{\phi \in \Phi_f} \phi(w) \tag{67}$$

where

$$\Psi_f = \{\psi : \psi \text{ is concave in } [0,1]^V \text{ and } \psi(1_X) \geq f(X), \forall X \subseteq V\} \tag{68}$$

- We can express this using the generalized submodular upper polyhedron:

Lemma 17

The concave extension of any set function f can be expressed as:

$$\hat{f}(w) = \min_{(y,c) \in \mathcal{P}^f_{gen}} \langle y, w \rangle + c, \forall w \in [0,1]^{|V|} \tag{69}$$

- Unfortunately, the concave extension is NP hard to evaluate and optimize! ☹

An equivalent distributional characterization of the concave extension

- An equivalent characterization of the concave extension is from the distributional perspective.

An equivalent distributional characterization of the concave extension

- An equivalent characterization of the concave extension is from the distributional perspective.
- Denote Λ_w as the set:

$$\Lambda_w = \{\lambda_S, S \subseteq V : \sum_{S \subseteq V} \lambda_S 1_S = w, \sum_{S \subseteq V} \lambda_S = 1 \tag{70}$$

The concave extension above can also be represented as:

$$\hat{f}(w) = \max_{\lambda \in \Lambda_w} \lambda_S f(S) \tag{71}$$

An equivalent distributional characterization of the concave extension

- An equivalent characterization of the concave extension is from the distributional perspective.
- Denote Λ_w as the set:

$$\Lambda_w = \{\lambda_S, S \subseteq V : \sum_{S \subseteq V} \lambda_S 1_S = w, \sum_{S \subseteq V} \lambda_S = 1 \tag{70}$$

The concave extension above can also be represented as:

$$\hat{f}(w) = \max_{\lambda \in \Lambda_w} \lambda_S f(S) \tag{71}$$

- This is still NP hard to evaluate and optimize! ☹

Multilinear Extension

- A near-concave extension, used practically in algorithms is the multilinear extension!

$$\tilde{f}(x) = \sum_{X \subseteq V} f(X) \prod_{i \in X} x_i \prod_{i \notin X} (1 - x_i) \tag{72}$$

Multilinear Extension

- A near-concave extension, used practically in algorithms is the multilinear extension!

$$\tilde{f}(x) = \sum_{X \subseteq V} f(X) \prod_{i \in X} x_i \prod_{i \notin X} (1 - x_i) \tag{72}$$

- Can be seen to be related to the distributional perspective, since it is defined via a specific distribution p_X.

Multilinear Extension

- A near-concave extension, used practically in algorithms is the multilinear extension!

$$\tilde{f}(x) = \sum_{X \subseteq V} f(X) \prod_{i \in X} x_i \prod_{i \notin X} (1 - x_i) \quad (72)$$

- Can be seen to be related to the distributional perspective, since it is defined via a specific distribution p_X.
- Requires an exponential sum ☹, but can be approximated through sampling (Vondrak, 2007).

Multilinear Extension

- A near-concave extension, used practically in algorithms is the multilinear extension!

$$\tilde{f}(x) = \sum_{X \subseteq V} f(X) \prod_{i \in X} x_i \prod_{i \notin X} (1 - x_i) \tag{72}$$

- Can be seen to be related to the distributional perspective, since it is defined via a specific distribution p_X.
- Requires an exponential sum ☹, but can be approximated through sampling (Vondrak, 2007).
- For subclasses of submodular functions, one can compute the exact multilinear extension! (I-Jegelka-Bilmes, 2014) ☺

Outline

Some Optimality Conditions

Lemma 18

For a submodular function f, a set A is a maximizer of f, if $\mathbf{0} \in \partial^f(A)$.

Proposition 19

For a submodular function f, if $0 \in \partial^f_{(1,1)}(A)$ then A is a local maxima of f. By (Feige, Mirrokni, Vondrák,'07), this can also offer us a solution $S = \operatorname{argmax}_{X \in \{A, V \setminus A\}} f(X)$ with $f(S) \geq \frac{1}{3} OPT$.

Lemma 20

Given A s.t. $\mathbf{0} \in \partial^f_{i,(1,2)}(A)$, then A is the global maxima f.

Proof.

Immediate from the fact that $\partial^f_{i,(1,2)}(A) \subseteq \partial^f(A)$. □

- For various varieties of superdifferentials, other optimality conditions possible, including matroid constraints, etc.

Concave Discrete Separation Theorem

Lemma 21

Given submodular f and supermodular g, with $f(X) \leq g(X), \forall X \subseteq V$, and $f(\emptyset) = g(\emptyset)$ or $f(V) = g(V)$. There exists modular h such that $f(X) \leq h(X) \leq g(X), \forall X \subseteq V$. When f and g are also integral, there exists an integral h satisfying the above.

Proof.

Assume $f(\emptyset) = g(\emptyset)$. Then the following chain of inequalities hold:

$$f(X) \leq f(\emptyset) + \sum_{j \in X} f(j|\emptyset) \leq g(\emptyset) + \sum_{j \in X} g(j|\emptyset) \leq g(X) \tag{73}$$

Since $f(j|\emptyset) = f(j) - f(\emptyset) \leq g(j) - g(\emptyset) = g(j|\emptyset)$. The rest of the inequalities follow from submodularity (and supermodularity) of f (and g). $f(V) = g(V)$ holds analogously via functions $f(V \backslash X)$ and $g(V \backslash X)$. □

Fenchel from Concave Perspective

Lemma 22

Given submodular f and supermodular g such that the discrete separation theorem holds,

$$\max_{X \subseteq V} f(X) - g(X) = \min_{x} g_*(x) - f_*(x) \tag{74}$$

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

Proof.

The proof follows immediately from Theorem 4 of Fujishige&Narayanan'05, stating that Fenchel duality follows from discrete separation, if the same conditions hold. □

Murota'03 proved that for $M^\natural$-concave and $M^\natural$-convex functions respectively, the above form of Fenchel duality always holds.