Submodular Functions, Optimization, and Applications to Machine Learning

— Spring Quarter, Lecture 4 —

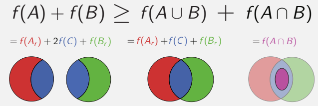
http://www.ee.washington.edu/people/faculty/bilmes/classes/ee596b_spring_2016/

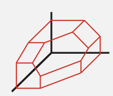
Prof. Jeff Bilmes

University of Washington, Seattle
Department of Electrical Engineering
http://melodi.ee.washington.edu/~bilmes

Apr 6th, 2016

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





Cumulative Outstanding Reading

- Read chapter 1 from Fujishige's book.

Announcements, Assignments, and Reminders

- Homework 1 is now available at our assignment dropbox (`https://canvas.uw.edu/courses/1039754/assignments`), due (electronically) Friday at 11:55pm.
- Weekly Office Hours: Mondays, 3:30-4:30, or by skype or google hangout (set up meeting via our our discussion board (`https://canvas.uw.edu/courses/1039754/discussion_topics`)).

Class Road Map - IT-I

- L1(3/28): Motivation, Applications, & Basic Definitions
- L2(3/30): Machine Learning Apps (diversity, complexity, parameter, learning target, surrogate).
- L3(4/4): Info theory exs, more apps, definitions, graph/combinatorial examples, matrix rank example, visualization
- L4(4/6): Graph and Combinatorial Examples, matrix rank, Venn diagrams, examples of proofs of submodularity, some useful properties
- L5(4/11):
- L6(4/13):
- L7(4/18):
- L8(4/20):
- L9(4/25):
- L10(4/27):
- L11(5/2):
- L12(5/4):
- L13(5/9):
- L14(5/11):
- L15(5/16):
- L16(5/18):
- L17(5/23):
- L18(5/25):
- L19(6/1):
- L20(6/6): Final Presentations maximization.

Finals Week: June 6th-10th, 2016.

Monge Matrices

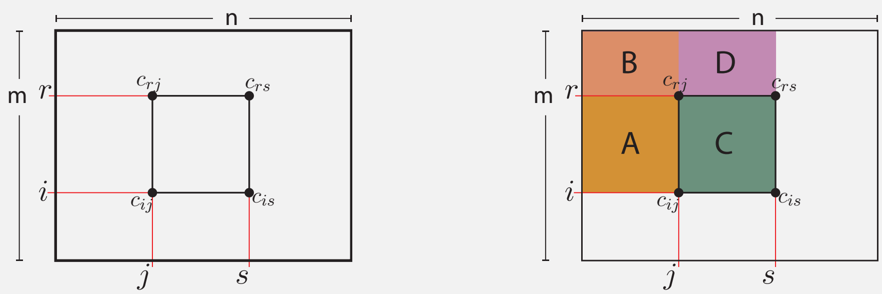
- $m \times n$ matrices $C = [c_{ij}]_{ij}$ are called Monge matrices if they satisfy the Monge property, namely:

$$c_{ij} + c_{rs} \leq c_{is} + c_{rj} \tag{4.15}$$

  for all $1 \leq i < r \leq m$ and $1 \leq j < s \leq n$.
- Equivalently, for all $1 \leq i, r \leq m$, $1 \leq j, s \leq n$,

$$c_{\min(i,r),\min(j,s)} + c_{\max(i,r),\max(j,s)} \leq c_{is} + c_{rj} \tag{4.16}$$

- Consider four elements of the $m \times n$ matrix:

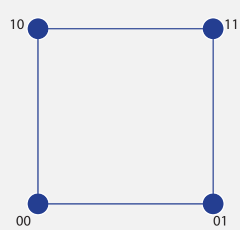


$c_{ij} = A + B$, $c_{rj} = B$, $c_{rs} = B + D$, $c_{is} = A + B + C + D$.

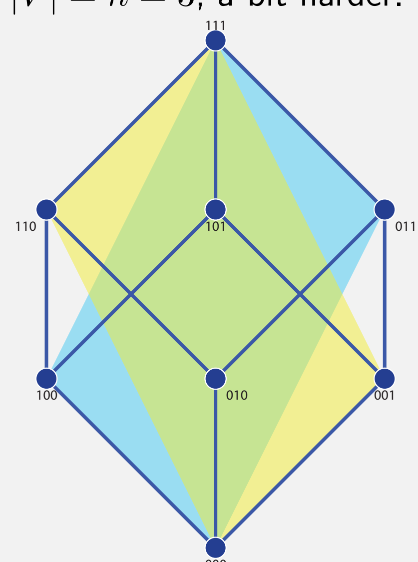
Submodular on Hypercube Vertices

- Test submodularity via values on verticies of hypercube.

Example: with $|V| = n = 2$, this is easy:



With $|V| = n = 3$, a bit harder.



How many inequalities?

Subadditive Definitions

Definition 4.2.1 (subadditive)

A function $f : 2^V \to \mathbb{R}$ is subadditive if for any $A, B \subseteq V$, we have that:

$$f(A) + f(B) \geq f(A \cup B) \tag{4.21}$$

This means that the "whole" is less than the sum of the parts.

Superadditive Definitions

Definition 4.2.1 (superadditive)

A function $f : 2^V \to \mathbb{R}$ is superadditive if for any $A, B \subseteq V$, we have that:

$$f(A) + f(B) \leq f(A \cup B) \tag{4.21}$$

- This means that the "whole" is greater than the sum of the parts.
- In general, submodular and subadditive (and supermodular and superadditive) are different properties.
- Ex: Let $0 < k < |V|$, and consider $f : 2^V \to \mathbb{R}_+$ where:

$$f(A) = \begin{cases} 1 & \text{if } |A| \leq k \\ 0 & \text{else} \end{cases} \tag{4.22}$$

- This function is subadditive but not submodular.

Modular Definitions

Definition 4.2.1 (modular)

A function that is both submodular and supermodular is called **modular**

If f is a modular function, than for any $A, B \subseteq V$, we have

$$f(A) + f(B) = f(A \cap B) + f(A \cup B) \tag{4.21}$$

In modular functions, elements do not interact (or cooperate, or compete, or influence each other), and have value based only on singleton values.

Proposition 4.2.2

If f is modular, it may be written as

$$f(A) = f(\emptyset) + \sum_{a \in A} \Big(f(\{a\}) - f(\emptyset) \Big) = c + \sum_{a \in A} f'(a) \tag{4.22}$$

which has only $|V| + 1$ parameters.

Complement function

Given a function $f : 2^V \to \mathbb{R}$, we can find a complement function $\bar{f} : 2^V \to \mathbb{R}$ as $\bar{f}(A) = f(V \setminus A)$ for any A.

Proposition 4.2.1

$\bar{f}$ is submodular iff f is submodular.

Proof.

$$\bar{f}(A) + \bar{f}(B) \geq \bar{f}(A \cup B) + \bar{f}(A \cap B) \tag{4.26}$$

follows from

$$f(V \setminus A) + f(V \setminus B) \geq f(V \setminus (A \cup B)) + f(V \setminus (A \cap B)) \tag{4.27}$$

which is true because $V \setminus (A \cup B) = (V \setminus A) \cap (V \setminus B)$ and $V \setminus (A \cap B) = (V \setminus A) \cup (V \setminus B)$ (De Morgan's laws for sets). $\square$

Other graph functions that are submodular/supermodular

These come from Narayanan's book 1997. Let G be an undirected graph.

- Let $V(X)$ be the vertices adjacent to some edge in $X \subseteq E(G)$, then $|V(X)|$ (the vertex function) is submodular.
- Let $E(S)$ be the edges with both vertices in $S \subseteq V(G)$. Then $|E(S)|$ (the interior edge function) is supermodular.
- Let $I(S)$ be the edges with at least one vertex in $S \subseteq V(G)$. Then $|I(S)|$ (the incidence function) is submodular.
- Recall $|\delta(S)|$, is the set size of edges with exactly one vertex in $S \subseteq V(G)$ is submodular (cut size function). Thus, we have $I(S) = E(S) \cup \delta(S)$ and $E(S) \cap \delta(S) = \emptyset$, and thus that $|I(S)| = |E(S)| + |\delta(S)|$. So we can get a submodular function by summing a submodular and a supermodular function. If you had to guess, is this always the case?
- Consider $f(A) = |\delta^+(A)| - |\delta^+(V \setminus A)|$. Guess, submodular, supermodular, modular, or neither? Exercise: determine which one and prove it.

Number of connected components in a graph via edges

- Recall, $f : 2^V \to \mathbb{R}$ is submodular, then so is $\bar{f} : 2^V \to \mathbb{R}$ defined as $\bar{f}(S) = f(V \setminus S)$.
- Hence, if $f : 2^V \to \mathbb{R}$ is supermodular, then so is $\bar{f} : 2^V \to \mathbb{R}$ defined as $\bar{f}(S) = f(V \setminus S)$.
- Given a graph $G = (V, E)$, for each $A \subseteq E(G)$, let $c(A)$ denote the number of connected components of the (spanning) subgraph $(V(G), A)$, with $c : 2^E \to \mathbb{R}_+$.
- $c(A)$ is monotone non-increasing, $c(A + a) - c(A) \leq 0$.
- Then $c(A)$ is supermodular, i.e.,
$$c(A + a) - c(A) \leq c(B + a) - c(B) \tag{4.40}$$
with $A \subseteq B \subseteq E \setminus \{a\}$.
- Intuition: an edge is "more" (no less) able to bridge separate components (and reduce the number of conected components) when edge is added in a smaller context than when added in a larger context.
- $\bar{c}(A) = c(E \setminus A)$ is the number of connected components in G when we remove A, so is also supermodular, but monotone non-decreasing.

Graph Strength

- So $\bar{c}(A) = c(E \setminus A)$ is the number of connected components in G when we remove A, is supermodular.
- Maximizing $\bar{c}(A)$ might seem as a goal for a network attacker — many connected components means that many points in the network have lost connectivity to many other points (unprotected network).
- If we can remove a small set A and shatter the graph into many connected components, then the graph is weak.
- An attacker wishes to choose a small number of edges (since it is cheap) to shatter the graph into as many components as possible.
- Let $G = (V, E, w)$ with $w : E \to \mathbb{R}+$ be a weighted graph with non-negative weights.
- For $(u, v) = e \in E$, let $w(e)$ be a measure of the strength of the connection between vertices u and v (strength meaning the difficulty of cutting the edge e).

Graph Strength

- Then $w(A)$ for $A \subseteq E$ is a modular function

$$w(A) = \sum_{e \in A} w_e \tag{4.1}$$

so that $w(E(G[S]))$ is the "internal strength" of the vertex set S.
- Suppose removing A shatters G into a graph with $\bar{c}(A) > 1$ components — then $w(A)/(\bar{c}(A) - 1)$ is like the "effort per achieved/additional component" for a network attacker.
- A form of graph strength can then be defined as the following:

$$\textit{strength}(G, w) = \min_{A \subseteq E(G): \bar{c}(A) > 1} \frac{w(A)}{\bar{c}(A) - 1} \tag{4.2}$$

- Graph strength is like the minimum effort per component. An attacker would use the argument of the min to choose which edges to attack. A network designer would maximize, over G and/or w, the graph strength, *strength*(G, w).
- Since submodularity, problems have strongly-poly-time solutions.

Submodularity, Quadratic Structures, and Cuts

Lemma 4.3.1

Let $\mathbf{M} \in \mathbb{R}^{n \times n}$ be a symmetric matrix and $m \in \mathbb{R}^n$ be a vector. Then $f : 2^V \to \mathbb{R}$ defined as

$$f(X) = m^\intercal \mathbf{1}_X + \frac{1}{2}\mathbf{1}_X^\intercal \mathbf{M} \mathbf{1}_X \tag{4.3}$$

is submodular iff the off-diagonal elements of M are non-positive.

Proof.

- Given a complete graph $G = (V, E)$, recall that $E(X)$ is the edge set with both vertices in $X \subseteq V(G)$, and that $|E(X)|$ is supermodular.
- Non-negative modular weights $w^+ : E \to \mathbb{R}_+$, $w(E(X))$ is also supermodular, so $-w(E(X))$ (non-positive modular) is submodular.
- f is a modular function $m^\intercal \mathbf{1}_A = m(A)$ added to a weighted submodular function, hence f is submodular.

Submodularity, Quadratic Structures, and Cuts

Proof of Lemma 4.3.1 cont.

- Conversely, suppose f is submodular.
- Then $\forall u, v \in V$, $f(\{u\}) + f(\{v\}) \geq f(\{u, v\}) + f(\emptyset)$ while $f(\emptyset) = 0$.
- This requires:

$$\begin{aligned} 0 &\leq f(\{u\}) + f(\{v\}) - f(\{u, v\}) && (4.4)\\ &= m(u) + \frac{1}{2}M_{u,u} + m(v) + \frac{1}{2}M_{v,v} && (4.5)\\ &\quad - \Big(m(u) + m(v) + \frac{1}{2}M_{u,u} + M_{u,v} + \frac{1}{2}M_{v,v}\Big) && (4.6)\\ &= -M_{u,v} && (4.7) \end{aligned}$$

So that $\forall u, v \in V$, $M_{u,v} \leq 0$.

$\square$

SET COVER and MAXIMUM COVERAGE

just Special cases of Submodular Optimization

- We are given a finite set V of n elements and a set of subsets $\mathcal{V} = \{V_1, V_2, \ldots, V_m\}$ of m subsets of V, so that $V_i \subseteq V$ and $\bigcup_i V_i = V$.
- The goal of minimum SET COVER is to choose the smallest subset $A \subseteq [m] \triangleq \{1, \ldots, m\}$ such that $\bigcup_{a \in A} V_a = V$.
- Maximum k cover: The goal in MAXIMUM COVERAGE is, given an integer $k \le m$, select k subsets, say $\{a_1, a_2, \ldots, a_k\}$ with $a_i \in [m]$ such that $|\bigcup_{i=1}^k V_{a_i}|$ is maximized.
- $f : 2^{[m]} \to \mathbb{Z}_+$ where for $A \subseteq [m]$, $f(A) = |\bigcup_{a \in A} V_a|$ is the set cover function and is submodular.
- Both SET COVER and MAXIMUM COVERAGE are well known to be NP-hard, but have a fast greedy approximation algorithm, and hence are instances of submodular optimization.

Vertex and Edge Covers

Also instances of submodular optimization

Definition 4.3.2 (vertex cover)

A *vertex cover* (a "vertex-based cover of edges") in graph $G = (V, E)$ is a set $S \subseteq V(G)$ of vertices such that every edge in G is incident to at least one vertex in S.

- Let $I(S)$ be the number of edges incident to vertex set S. Then we wish to find the smallest set $S \subseteq V$ subject to $I(S) = |E|$.

Definition 4.3.3 (edge cover)

A *edge cover* (an "edge-based cover of vertices") in graph $G = (V, E)$ is a set $F \subseteq E(G)$ of edges such that every vertex in G is incident to at least one edge in F.

- Let $|V|(F)$ be the number of vertices incident to edge set F. Then we wish to find the smallest set $F \subseteq E$ subject to $|V|(F) = |V|$.

Graph Cut Problems
Also submodular optimization

- MINIMUM CUT: Given a graph $G = (V, E)$, find a set of vertices $S \subseteq V$ that minimize the cut (set of edges) between S and $V \setminus S$.
- MAXIMUM CUT: Given a graph $G = (V, E)$, find a set of vertices $S \subseteq V$ that minimize the cut (set of edges) between S and $V \setminus S$.
- Let $\delta : 2^V \to \mathbb{R}_+$ be the cut function, namely for any given set of nodes $X \subseteq V$, $\delta(X)$ measures the number of edges between nodes X and $V \setminus X$, or $\delta(x) = E(X, V \setminus X)$.
- Weighted versions, where rather than count, we sum the (non-negative) weights of the edges of a cut, $f(X) = w(\delta(X))$.
- Hence, MINIMUM CUT and MAXIMUM CUT are also special cases of submodular optimization.

Matrix Rank functions

- Let V, with $|V| = m$ be an index set of a set of vectors in $\mathbb{R}^n$ for some n (unrelated to m).
- For a given set $\{v, v_1, v_2, \ldots, v_k\}$, it might or might not be possible to find $(\alpha_i)_i$ such that:

$$x_v = \sum_{i=1}^{k} \alpha_i x_{v_i} \tag{4.8}$$

  If not, then x_v is linearly independent of $x_{v_1}, \ldots, x_{v_k}$.
- Let $r(S)$ for $S \subseteq V$ be the rank of the set of vectors S. Then $r(\cdot)$ is a submodular function, and in fact is called a matric matroid rank function.

Example: Rank function of a matrix

- Given $n \times m$ matrix $\mathbf{X} = (x_1, x_2, \ldots, x_m)$ with $x_i \in \mathbb{R}^n$ for all i. There are m length-n column vectors $\{x_i\}_i$
- Let $V = \{1, 2, \ldots, m\}$ be the set of column vector indices.
- For any $A \subseteq V$, let $r(A)$ be the rank of the column vectors indexed by A.
- $r(A)$ is the dimensionality of the vector space spanned by the set of vectors $\{x_a\}_{a \in A}$.
- Thus, $r(V)$ is the rank of the matrix $\mathbf{X}$.

▸ Skip matrix rank example

Graph & Combinatorial Examples | Matrix Rank | Examples and Properties | Other Submodular Defs.

Example: Rank function of a matrix

Consider the following 4×8 matrix, so $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$.

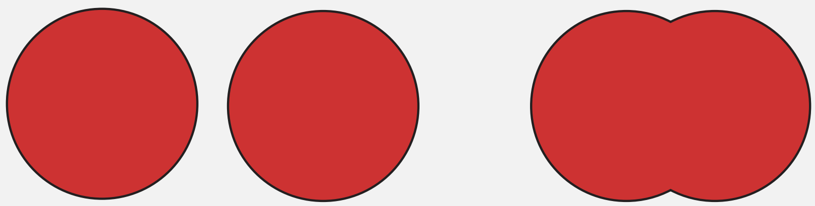
$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array}\begin{array}{c} \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{matrix} \\ \begin{pmatrix} 0 & 2 & 2 & 3 & 0 & 1 & 3 & 1 \\ 0 & 3 & 0 & 4 & 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 & 3 & 0 & 0 & 5 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 5 \end{pmatrix} \end{array} = \begin{array}{c} \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{matrix} \\ \begin{pmatrix} | & | & | & | & | & | & | & | \\ x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 \\ | & | & | & | & | & | & | & | \end{pmatrix} \end{array}$$

- Let $A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$, $C = \{6, 7\}$, $A_r = \{1\}$, $B_r = \{5\}$.
- Then $r(A) = 3$, $r(B) = 3$, $r(C) = 2$.
- $r(A \cup C) = 3$, $r(B \cup C) = 3$.
- $r(A \cup A_r) = 3$, $r(B \cup B_r) = 3$, $r(A \cup B_r) = 4$, $r(B \cup A_r) = 4$.
- $r(A \cup B) = 4$, $r(A \cap B) = 1 < r(C) = 2$.
- $6 = r(A) + r(B) > r(A \cup B) + r(A \cap B) = 5$

Rank function of a matrix

- Let $A, B \subseteq V$ be two subsets of column indices.
- The rank of the two sets unioned together $A \cup B$ is no more than the sum of the two individual ranks.
- In a Venn diagram, let area correspond to dimensions spanned by vectors indexed by a set. Hence, $r(A)$ can be viewed as an area.

$$r(A) \quad + \quad r(B) \qquad \geq \qquad r(A \cup B)$$



- If some of the dimensions spanned by A overlap some of the dimensions spanned by B (i.e., if $\exists$ common span), then that area is counted twice in $r(A) + r(B)$, so the inequality will be strict.
- Any function where the above inequality is true for all $A, B \subseteq V$ is called subadditive.

Rank functions of a matrix

- Vectors A and B have a (possibly empty) common span and two (possibly empty) non-common residual spans.
- Let C index vectors spanning dimensions common to A and B.
- Let A_r index vectors spanning dimensions spanned by A but not B.
- Let B_r index vectors spanning dimensions spanned by B but not A.
- Then, $r(A) = r(C) + r(A_r)$
- Similarly, $r(B) = r(C) + r(B_r)$.
- Then $r(A) + r(B)$ counts the dimensions spanned by C twice, i.e.,

$$r(A) + r(B) = r(A_r) + 2r(C) + r(B_r). \tag{4.9}$$

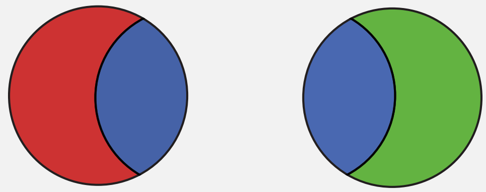
- But $r(A \cup B)$ counts the dimensions spanned by C only once.

$$r(A \cup B) = r(A_r) + r(C) + r(B_r) \tag{4.10}$$

Rank functions of a matrix

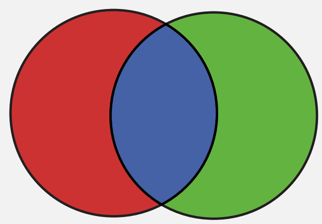
- Then $r(A) + r(B)$ counts the dimensions spanned by C twice, i.e.,

$$r(A) + r(B) = r(A_r) + 2r(C) + r(B_r)$$



- But $r(A \cup B)$ counts the dimensions spanned by C only once.

$$r(A \cup B) = r(A_r) + r(C) + r(B_r)$$

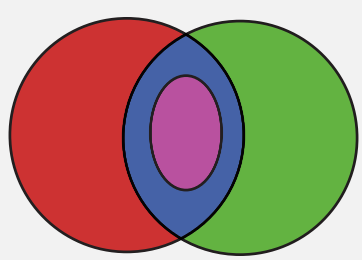


- Thus, we have subadditivity: $r(A) + r(B) \geq r(A \cup B)$. Can we add more to the r.h.s. and still have an inequality? Yes.

Rank function of a matrix

- Note, $r(A \cap B) \leq r(C)$. Why? Vectors indexed by $A \cap B$ (i.e., the common index set) span no more than the dimensions commonly spanned by A and B (namely, those spanned by the professed C).

$$r(C) \geq r(A \cap B)$$

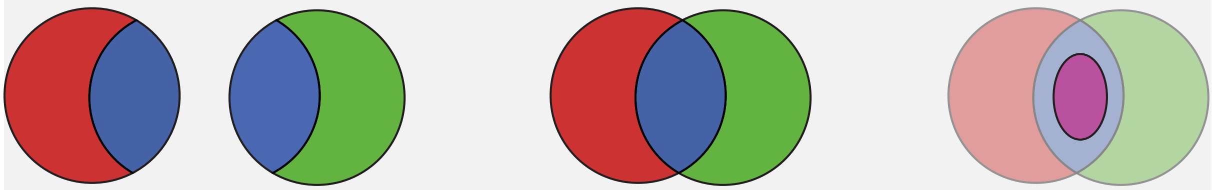


In short:

- Common span (blue) is "more" (no less) than span of common index (magenta).
- More generally, common information (blue) is "more" (no less) than information within common index (magenta).

The Venn and Art of Submodularity

$$\underbrace{r(A) + r(B)}_{= r(A_r) + 2r(C) + r(B_r)} \geq \underbrace{r(A \cup B)}_{= r(A_r) + r(C) + r(B_r)} + \underbrace{r(A \cap B)}_{= r(A \cap B)}$$



Polymatroid rank function

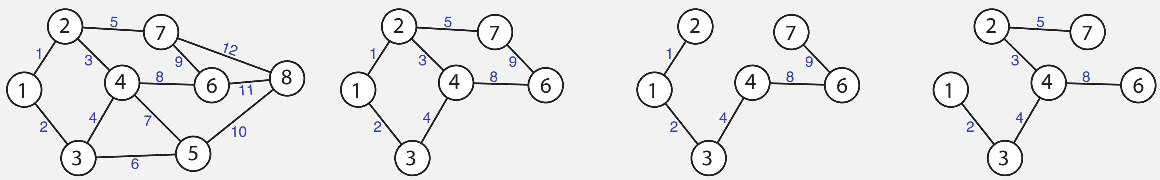
- Let S be a set of subspaces of a linear space (i.e., each $s \in S$ is a subspace of dimension ≥ 1).
- For each $X \subseteq S$, let $f(X)$ denote the dimensionality of the linear subspace spanned by the subspaces in X.
- We can think of S as a set of sets of vectors from the matrix rank example, and for each $s \in S$, let X_s being a set of vector indices.
- Then, defining $f : 2^S \to \mathbb{R}_+$ as follows,

$$f(X) = r(\cup_{s \in S} X_s) \tag{4.11}$$

  we have that f is submodular, and is known to be a polymatroid rank function.
- In general (as we will see) polymatroid rank functions are submodular, normalized $f(\emptyset) = 0$, and monotone non-decreasing ($f(A) \leq f(B)$ whenever $A \subseteq B$).

Spanning trees

- Let E be a set of edges of some graph $G = (V, E)$, and let $r(S)$ for $S \subseteq E$ be the maximum size (in terms of number of edges) spanning forest in the vertex-induced graph, induced by vertices incident to edges S.
- Example: Given $G = (V, E)$, $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$, $E = \{1, 2, \ldots, 12\}$. $S = \{1, 2, 3, 4, 5, 8, 9\} \subset E$. Two spanning trees have the same edge count (the rank of S).



- Then $r(S)$ is submodular, and is another matrix rank function corresponding to the incidence matrix of the graph.

Submodular Polyhedra

- Submodular functions have associated polyhedra with nice properties: when a set of constraints in a linear program is a submodular polyhedron, a simple greedy algorithm can find the optimal solution even though the polyhedron is formed via an exponential number of constraints.

$$P_f = \left\{x \in \mathbb{R}^E : x(S) \le f(S), \forall S \subseteq E\right\} \tag{4.12}$$

$$P_f^+ = P_f \cap \left\{x \in \mathbb{R}^E : x \ge 0\right\} \tag{4.13}$$

$$B_f = P_f \cap \left\{x \in \mathbb{R}^E : x(E) = f(E)\right\} \tag{4.14}$$

- The linear programming problem is to, given $c \in \mathbb{R}^E$, compute:

$$\tilde{f}(c) \triangleq \max\left\{c^T x : x \in P_f\right\} \tag{4.15}$$

- This can be solved using the greedy algorithm! Moreover, $\tilde{f}(c)$ computed using greedy is convex if and only of f is submodular (we will go into this in some detail this quarter).

Summing Submodular Functions

Given E, let $f_1, f_2 : 2^E \to \mathbb{R}$ be two submodular functions. Then

$$f : 2^E \to \mathbb{R} \text{ with } f(A) = f_1(A) + f_2(A) \tag{4.16}$$

is submodular.This follows easily since

$$\begin{aligned} f(A) + f(B) &= f_1(A) + f_2(A) + f_1(B) + f_2(B) && (4.17)\\ &\geq f_1(A \cup B) + f_2(A \cup B) + f_1(A \cap B) + f_2(A \cap B) && (4.18)\\ &= f(A \cup B) + f(A \cap B). && (4.19) \end{aligned}$$

I.e., it holds for each component of f in each term in the inequality. In fact, any conic combination (i.e., non-negative linear combination) of submodular functions is submodular, as in $f(A) = \alpha_1 f_1(A) + \alpha_2 f_2(A)$ for $\alpha_1, \alpha_2 \geq 0$.

Summing Submodular and Modular Functions

Given E, let $f_1, m : 2^E \to \mathbb{R}$ be a submodular and a modular function. Then

$$f : 2^E \to \mathbb{R} \text{ with } f(A) = f_1(A) - m(A) \tag{4.20}$$

is submodular (as is $f(A) = f_1(A) + m(A)$). This follows easily since

$$\begin{aligned} f(A) + f(B) &= f_1(A) - m(A) + f_1(B) - m(B) && (4.21)\\ &\geq f_1(A \cup B) - m(A \cup B) + f_1(A \cap B) - m(A \cap B) && (4.22)\\ &= f(A \cup B) + f(A \cap B). && (4.23) \end{aligned}$$

That is, the modular component with $m(A) + m(B) = m(A \cup B) + m(A \cap B)$ never destroys the inequality. Note of course that if m is modular than so is $-m$.

Restricting Submodular Functions

Given E, let $f : 2^E \to \mathbb{R}$ be a submodular functions. And let $S \subseteq E$ be an arbitrary fixed set. Then

$$f' : 2^E \to \mathbb{R} \text{ with } f'(A) = f(A \cap S) \tag{4.24}$$

is submodular.

Proof.

Given $A \subseteq B \subseteq E \setminus v$, consider

$$f((A + v) \cap S) - f(A \cap S) \geq f((B + v) \cap S) - f(B \cap S) \tag{4.25}$$

If $v \notin S$, then both differences on each size are zero. If $v \in S$, then we can consider this

$$f(A' + v) - f(A') \geq f(B' + v) - f(B') \tag{4.26}$$

with $A' = A \cap S$ and $B' = B \cap S$. Since $A' \subseteq B'$, this holds due to submodularity of f. □

Summing Restricted Submodular Functions

Given V, let $f_1, f_2 : 2^V \to \mathbb{R}$ be two submodular functions and let S_1, S_2 be two arbitrary fixed sets. Then

$$f : 2^V \to \mathbb{R} \text{ with } f(A) = f_1(A \cap S_1) + f_2(A \cap S_2) \tag{4.27}$$

is submodular. This follows easily from the preceding two results.
Given V, let $\mathcal{C} = \{C_1, C_2, \ldots, C_k\}$ be a set of subsets of V, and for each $C \in \mathcal{C}$, let $f_C : 2^V \to \mathbb{R}$ be a submodular function. Then

$$f : 2^V \to \mathbb{R} \text{ with } f(A) = \sum_{C \in \mathcal{C}} f_C(A \cap C) \tag{4.28}$$

is submodular. This property is critical for image processing and graphical models. For example, let $\mathcal{C}$ be all pairs of the form $\{\{u, v\} : u, v \in V\}$, or let it be all pairs corresponding to the edges of some undirected graphical model. We plan to revisit this topic later in the term.

Max - normalized

Given V, let $c \in \mathbb{R}_+^V$ be a given fixed vector. Then $f : 2^V \to \mathbb{R}_+$, where

$$f(A) = \max_{j \in A} c_j \tag{4.29}$$

is submodular and normalized (we take $f(\emptyset) = 0$).

Proof.

Consider

$$\max_{j \in A} c_j + \max_{j \in B} c_j \geq \max_{j \in A \cup B} c_j + \max_{j \in A \cap B} c_j \tag{4.30}$$

which follows since we have that

$$\max(\max_{j \in A} c_j, \max_{j \in B} c_j) = \max_{j \in A \cup B} c_j \tag{4.31}$$

and

$$\min(\max_{j \in A} c_j, \max_{j \in B} c_j) \geq \max_{j \in A \cap B} c_j \tag{4.32}$$

□

Max

Given V, let $c \in \mathbb{R}^V$ be a given fixed vector (not necessarily non-negative). Then $f : 2^V \to \mathbb{R}$, where

$$f(A) = \max_{j \in A} c_j \tag{4.33}$$

is submodular, where we take $f(\emptyset) \leq \min_j c_j$ (so the function is not normalized).

Proof.

The proof is identical to the normalized case. □

Facility/Plant Location (uncapacitated) w. plant benefits

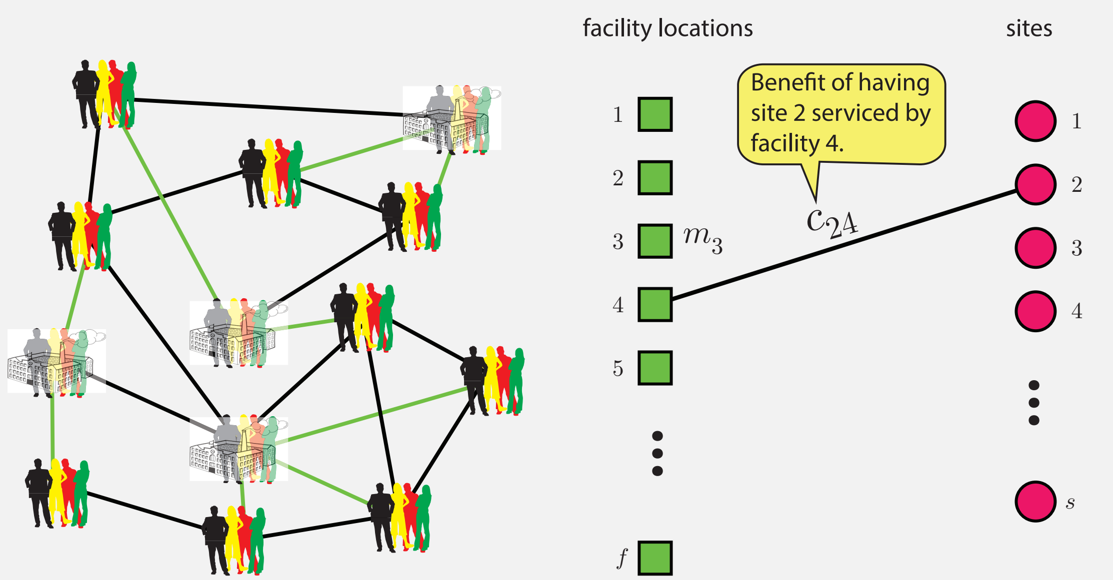
- Let $F = \{1, \dots, f\}$ be a set of possible factory/plant locations for facilities to be built.
- $S = \{1, \dots, s\}$ is a set of sites (e.g., cities, clients) needing service.
- Let c_{ij} be the "benefit" (e.g., $1/c_{ij}$ is the cost) of servicing site i with facility location j.
- Let m_j be the benefit (e.g., either $1/m_j$ is the cost or $-m_j$ is the cost) to build a plant at location j.
- Each site should be serviced by only one plant but no less than one.
- Define $f(A)$ as the "delivery benefit" plus "construction benefit" when the locations $A \subseteq F$ are to be constructed.
- We can define the (uncapacitated) facility location function

$$f(A) = \sum_{j \in A} m_j + \sum_{i \in F} \max_{j \in A} c_{ij}. \tag{4.34}$$

- Goal is to find a set A that maximizes $f(A)$ (the benefit) placing a bound on the number of plants A (e.g., $|A| \leq k$).

Facility/Plant Location (uncapacitated)

- Core problem in operations research, early motivation for submodularity.
- Goal: as efficiently as possible, place "facilities" (factories) at certain locations to satisfy sites (at all locations) having various demands.



Facility Location

Given V, E, let $c \in \mathbb{R}^{V \times E}$ be a given $|V| \times |E|$ matrix. Then

$$f : 2^E \to \mathbb{R}, \text{ where } f(A) = \sum_{i \in V} \max_{j \in A} c_{ij} \tag{4.35}$$

is submodular.

Proof.

We can write $f(A)$ as $f(A) = \sum_{i \in V} f_i(A)$ where $f_i(A) = \max_{j \in A} c_{ij}$ is submodular (max of a i^{th} row vector), so f can be written as a sum of submodular functions. □

Thus, the facility location function (which only adds a modular function to the above) is submodular.

Log Determinant

- Let $\mathbf{\Sigma}$ be an $n \times n$ positive definite matrix. Let $V = \{1, 2, \ldots, n\} \equiv [n]$ be an index set, and for $A \subseteq V$, let $\mathbf{\Sigma}_A$ be the (square) submatrix of $\mathbf{\Sigma}$ obtained by including only entries in the rows/columns given by A.
- We have that:

$$f(A) = \log \det(\mathbf{\Sigma}_A) \text{ is submodular.} \tag{4.36}$$

- The submodularity of the log determinant is crucial for determinantal point processes (DPPs) (defined later in the class).

Proof of submodularity of the logdet function.

Suppose $X \in \mathbf{R}^n$ is multivariate Gaussian random variable, that is

$$x \in p(x) = \frac{1}{\sqrt{|2\pi\mathbf{\Sigma}|}} \exp\left(-\frac{1}{2}(x-\mu)^T \mathbf{\Sigma}^{-1}(x-\mu)\right) \tag{4.37}$$

...

Log Determinant

...cont.

Then the (differential) entropy of the r.v. X is given by

$$h(X) = \log \sqrt{|2\pi e \mathbf{\Sigma}|} = \log \sqrt{(2\pi e)^n |\mathbf{\Sigma}|} \tag{4.38}$$

and in particular, for a variable subset A,

$$f(A) = h(X_A) = \log \sqrt{(2\pi e)^{|A|} |\mathbf{\Sigma}_A|} \tag{4.39}$$

Entropy is submodular (further conditioning reduces entropy), and moreover

$$f(A) = h(X_A) = m(A) + \frac{1}{2} \log |\mathbf{\Sigma}_A| \tag{4.40}$$

where $m(A)$ is a modular function. □

Note: still submodular in the semi-definite case as well.

Summary so far

- Summing: if $\alpha_i \geq 0$ and $f_i : 2^V \to \mathbb{R}$ is submodular, then so is $\sum_i \alpha_i f_i$.
- Restrictions: $f'(A) = f(A \cap S)$
- max: $f(A) = \max_{j \in A} c_j$ and facility location.
- Log determinant $f(A) = \log \det(\mathbf{\Sigma}_A)$

Concave over non-negative modular

Let $m \in \mathbb{R}_+^E$ be a non-negative modular function, and g a concave function over $\mathbb{R}$. Define $f : 2^E \to \mathbb{R}$ as

$$f(A) = g(m(A)) \tag{4.41}$$

then f is submodular.

Proof.

Given $A \subseteq B \subseteq E \setminus v$, we have $0 \leq a = m(A) \leq b = m(B)$, and $0 \leq c = m(v)$. For g concave, we have $g(a+c) - g(a) \geq g(b+c) - g(b)$, and thus

$$g(m(A) + m(v)) - g(m(A)) \geq g(m(B) + m(v)) - g(m(B)) \tag{4.42}$$

□

A form of converse is true as well.

Concave composed with non-negative modular

Theorem 4.5.1

Given a ground set V. The following two are equivalent:

1. *For all modular functions $m : 2^V \to \mathbb{R}_+$, then $f : 2^V \to \mathbb{R}$ defined as $f(A) = g(m(A))$ is submodular*
2. *$g : \mathbb{R}_+ \to \mathbb{R}$ is concave.*

- If g is non-decreasing concave, then f is polymatroidal.
- Sums of concave over modular functions are submodular

$$f(A) = \sum_{i=1}^{K} g_i(m_i(A)) \tag{4.43}$$

- Very large class of functions, including graph cut, bipartite neighborhoods, set cover (Stobbe & Krause 2011), and "feature-based submodular functions" (Wei, Iyer, & Bilmes 2014).
- However, Vondrak showed that a graphic matroid rank function over K_4 (we'll define this after we define matroids) are not members.

Monotonicity

Definition 4.5.2

A function $f : 2^V \to \mathbb{R}$ is monotone nondecreasing (resp. monotone increasing) if for all $A \subset B$, we have $f(A) \le f(B)$ (resp. $f(A) < f(B)$).

Definition 4.5.3

A function $f : 2^V \to \mathbb{R}$ is monotone nonincreasing (resp. monotone decreasing) if for all $A \subset B$, we have $f(A) \ge f(B)$ (resp. $f(A) > f(B)$).

Composition of non-decreasting submodular and non-decreasing concave

Theorem 4.5.4

Given two functions, one defined on sets

$$f : 2^V \to \mathbb{R} \tag{4.44}$$

and another continuous valued one:

$$g : \mathbb{R} \to \mathbb{R} \tag{4.45}$$

the composition formed as $h = g \circ f : 2^V \to \mathbb{R}$ (defined as $h(S) = g(f(S))$) is nondecreasing submodular, if g is non-decreasing concave and f is nondecreasing submodular.

Monotone difference of two functions

Let f and g both be submodular functions on subsets of V and let $(f-g)(\cdot)$ be either monotone increasing or monotone decreasing. Then $h : 2^V \to R$ defined by

$$h(A) = \min(f(A), g(A)) \tag{4.46}$$

is submodular.

Proof.

If $h(A)$ agrees with f on both X and Y (or g on both X and Y), and since

$$f(X) + f(Y) \geq f(X \cup Y) + f(X \cap Y) \tag{4.47}$$

$$g(X) + g(Y) \geq g(X \cup Y) + g(X \cap Y), \tag{4.48}$$

the result (Equation 4.46 being submodular) follows since

$$\begin{matrix} f(X) + f(Y) \\ g(X) + g(Y) \end{matrix} \geq \min(f(X \cup Y), g(X \cup Y)) + \min(f(X \cap Y), g(X \cap Y)) \tag{4.49}$$

Monotone difference of two functions

...cont.

Otherwise, w.l.o.g., $h(X) = f(X)$ and $h(Y) = g(Y)$, giving

$$h(X) + h(Y) = f(X) + g(Y) \geq f(X \cup Y) + f(X \cap Y) + g(Y) - f(Y) \tag{4.50}$$

Assume the case where $f - g$ is monotone increasing. Hence, $f(X \cup Y) + g(Y) - f(Y) \geq g(X \cup Y)$ giving

$$h(X) + h(Y) \geq g(X \cup Y) + f(X \cap Y) \geq h(X \cup Y) + h(X \cap Y) \tag{4.51}$$

$\square$

What is an easy way to prove the case where $f - g$ is monotone decreasing?

Saturation via the $\min(\cdot)$ function

Let $f : 2^V \to \mathbb{R}$ be an monotone increasing or decreasing submodular function and let k be a constant. Then the function $h : 2^V \to \mathbb{R}$ defined by

$$h(A) = \min(k, f(A)) \tag{4.52}$$

is submodular.

Proof.

For constant k, we have that $(f - k)$ is increasing (or decreasing) so this follows from the previous result. □

Note also, $g(a) = \min(k, a)$ for constant k is a non-decreasing concave function, so when f is monotone nondecreasing submodular, we can use the earlier result about composing a monotone concave function with a monotone submodular function to get a version of this.

More on Min - the saturate trick

- In general, the minimum of two submodular functions is not submodular (unlike concave functions, closed under min).
- However, when wishing to maximize two monotone non-decreasing submodular functions f, g, we can define function $h : 2^V \to \mathbb{R}$ as

$$h_\alpha(A) = \min(\alpha, f(A)) + \min(\alpha, g(A)) \tag{4.53}$$

  then h is submodular, and $h(A) \geq k$ if and only if both $f(A) \geq \alpha$ and $g(A) \geq \alpha$, for constant $\alpha \in \mathbb{R}$.
- This can be useful in many applications. An instance of a submodular surrogate (where we take a non-submodular problem and find a submodular one that can tell us something).

Arbitrary functions as difference between submodular funcs.

Given an arbitrary set function f, it can be expressed as a difference between two submodular functions: $f = g - h$ where both g and h are submodular.

Proof.

Let f be given and arbitrary, and define:

$$\alpha \triangleq \min_{X,Y} \Big(f(X) + f(Y) - f(X \cup Y) - f(X \cap Y) \Big) \tag{4.54}$$

If $\alpha \geq 0$ then f is submodular, so by assumption $\alpha < 0$. Now let h be an arbitrary strict submodular function and define

$$\beta \triangleq \min_{X,Y : X \not\subseteq Y, Y \not\subseteq X} \Big(h(X) + h(Y) - h(X \cup Y) - h(X \cap Y) \Big). \tag{4.55}$$

Strict means that $\beta > 0$. ...

Arbitrary functions as difference between submodular funcs.

...cont.

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

$$f'(A) = f(A) + \frac{|\alpha|}{\beta} h(A) \tag{4.56}$$

Then f' is submodular (why?), and $f = f'(A) - \frac{|\alpha|}{\beta} h(A)$, a difference between two submodular functions as desired.

$\square$

Gain

- We often wish to express the gain of an item $j \in V$ in context A, namely $f(A \cup \{j\}) - f(A)$.
- This is called the gain and is used so often, there are equally as many ways to notate this. I.e., you might see:

$$f(A \cup \{j\}) - f(A) \triangleq \rho_j(A) \tag{4.57}$$
$$\triangleq \rho_A(j) \tag{4.58}$$
$$\triangleq \nabla_j f(A) \tag{4.59}$$
$$\triangleq f(\{j\}|A) \tag{4.60}$$
$$\triangleq f(j|A) \tag{4.61}$$

- We'll use $f(j|A)$.
- Submodularity's diminishing returns definition can be stated as saying that $f(j|A)$ is a monotone non-increasing function of A, since $f(j|A) \geq f(j|B)$ whenever $A \subseteq B$ (conditioning reduces valuation).

Gain Notation

It will also be useful to extend this to sets.
Let A, B be any two sets. Then

$$f(A|B) \triangleq f(A \cup B) - f(B) \tag{4.62}$$

So when j is any singleton

$$f(j|B) = f(\{j\}|B) = f(\{j\} \cup B) - f(B) \tag{4.63}$$

Note that this is inspired from information theory and the notation used for conditional entropy $H(X_A|X_B) = H(X_A, X_B) - H(X_B)$.

Arbitrary function as difference between two polymatroids

- Any normalized submodular function g can be represented as a sum of a polymatroid (normalized monotone non-decreasing submodular) function $\bar{g}$ and a modular function m_g.
- Given submodular $g : 2^V \to \mathbb{R}$, construct $\bar{g} : 2^V \to \mathbb{R}$ as $\bar{g}(A) = g(A) - \sum_{a \in A} g(a|V \setminus \{a\})$. Let $m_g(A) \triangleq \sum_{a \in A} g(a|V \setminus \{a\})$
- Then, given arbitrary $f = g - h$ where g and h are normalized submodular,

$$f = g - h = \bar{g} + m_g - (\bar{h} + m_h) \quad (4.64)$$
$$= \bar{g} - \bar{h} + (m_g - m_h) \quad (4.65)$$
$$= \bar{g} - \bar{h} + m_{g-h} \quad (4.66)$$
$$= \bar{g} + m^+_{g-h} - (\bar{h} + (-m_{g-h})^+) \quad (4.67)$$

where m^+ is the positive part of modular function m. That is, $m^+(A) = \sum_{a \in A} m(a)\mathbf{1}(m(a) > 0)$.
- But both $g + m^+_{g-h}$ and $\bar{h} + (-m_{g-h})^+$ are polymatroid functions.
- Thus, any function can be expressed as a difference between two, not only submodular (DS), but polymatroid functions.

Two Equivalent Submodular Definitions

Definition 4.6.1 (submodular concave)

A function $f : 2^V \to \mathbb{R}$ is submodular if for any $A, B \subseteq V$, we have that:

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B) \quad (4.8)$$

An alternate and (as we will soon see) equivalent definition is:

Definition 4.6.2 (diminishing returns)

A function $f : 2^V \to \mathbb{R}$ is submodular if for any $A \subseteq B \subset V$, and $v \in V \setminus B$, we have that:

$$f(A \cup \{v\}) - f(A) \geq f(B \cup \{v\}) - f(B) \quad (4.9)$$

The incremental "value", "gain", or "cost" of v decreases (diminishes) as the context in which v is considered grows from A to B.

Submodular Definition: Group Diminishing Returns

An alternate and equivalent definition is:

Definition 4.6.1 (group diminishing returns)

A function $f : 2^V \to \mathbb{R}$ is submodular if for any $A \subseteq B \subset V$, and $C \subseteq V \setminus B$, we have that:

$$f(A \cup C) - f(A) \geq f(B \cup C) - f(B) \tag{4.68}$$

This means that the incremental "value" or "gain" of set C decreases as the context in which C is considered grows from A to B (diminishing returns)

Submodular Definition Basic Equivalencies

We want to show that Submodular Concave (Definition 4.6.1), Diminishing Returns (Definition 4.6.2), and Group Diminishing Returns (Definition 4.6.1) are identical. We will show that:

- Submodular Concave $\Rightarrow$ Diminishing Returns
- Diminishing Returns $\Rightarrow$ Group Diminishing Returns
- Group Diminishing Returns $\Rightarrow$ Submodular Concave

Submodular Concave $\Rightarrow$ Diminishing Returns

$f(S) + f(T) \geq f(S \cup T) + f(S \cap T) \Rightarrow f(v|A) \geq f(v|B), A \subseteq B \subseteq V \setminus v.$

- Assume Submodular concave, so $\forall S, T$ we have $f(S) + f(T) \geq f(S \cup T) + f(S \cap T)$.
- Given A, B and $v \in V$ such that: $A \subseteq B \subseteq V \setminus \{v\}$, we have from submodular concave that:

$$f(A + v) + f(B) \geq f(B + v) + f(A) \tag{4.69}$$

- Rearranging, we have

$$f(A + v) - f(A) \geq f(B + v) - f(B) \tag{4.70}$$

□

Diminishing Returns $\Rightarrow$ Group Diminishing Returns

$f(v|S) \geq f(v|T), S \subseteq T \subseteq V \setminus v \Rightarrow f(C|A) \geq f(C|B), A \subseteq B \subseteq V \setminus C.$

Let $C = \{c_1, c_2, \ldots, c_k\}$. Then diminishing returns implies

$$\begin{aligned}
& f(A \cup C) - f(A) && (4.71)\\
&= f(A \cup C) - \sum_{i=1}^{k-1} \Big(f(A \cup \{c_1, \ldots, c_i\}) - f(A \cup \{c_1, \ldots, c_i\}) \Big) - f(A) && (4.72)\\
&= \sum_{i=1}^{k} \Big(f(A \cup \{c_1 \ldots c_i\}) - f(A \cup \{c_1 \ldots c_{i-1}\}) \Big) && (4.73)\\
&\geq \sum_{i=1}^{k} \Big(f(B \cup \{c_1 \ldots c_i\}) - f(B \cup \{c_1 \ldots c_{i-1}\}) \Big) && (4.74)\\
&= f(B \cup C) - \sum_{i=1}^{k-1} \Big(f(B \cup \{c_1, \ldots, c_i\}) - f(B \cup \{c_1, \ldots, c_i\}) \Big) - f(B) && (4.75)\\
&= f(B \cup C) - f(B) && (4.76)
\end{aligned}$$

□

Group Diminishing Returns $\Rightarrow$ Submodular Concave

$f(U|S) \geq f(U|T), S \subseteq T \subseteq V \setminus U \Rightarrow f(A) + f(B) \geq f(A \cup B) + f(A \cap B)$.

Assume group diminishing returns. Assume $A \neq B$ otherwise trivial. Define $A' = A \cap B$, $C = A \setminus B$, and $B' = B$. Then since $A' \subseteq B'$,

$$f(A' + C) - f(A') \geq f(B' + C) - f(B') \tag{4.77}$$

giving

$$f(A' + C) + f(B') \geq f(B' + C) + f(A') \tag{4.78}$$

or

$$f(A \cap B + A \setminus B) + f(B) \geq f(B + A \setminus B) + f(A \cap B) \tag{4.79}$$

which is the same as the submodular concave condition

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B) \tag{4.80}$$

Submodular Definition: Four Points

Definition 4.6.2 ("singleton", or "four points")

A function $f : 2^V \to \mathbb{R}$ is submodular iff for any $A \subset V$, and any $a, b \in V \setminus A$, we have that:

$$f(A \cup \{a\}) + f(A \cup \{b\}) \geq f(A \cup \{a, b\}) + f(A) \tag{4.81}$$

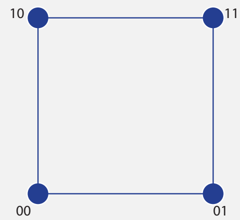
This follows immediately from diminishing returns. To achieve diminishing returns, assume $A \subset B$ with $B \setminus A = \{b_1, b_2, \ldots, b_k\}$. Then

$$\begin{aligned}
f(A + a) - f(A) &\geq f(A + b_1 + a) - f(A + b_1) && (4.82)\\
&\geq f(A + b_1 + b_2 + a) - f(A + b_1 + b_2) && (4.83)\\
&\geq \ldots && (4.84)\\
&\geq f(A + b_1 + \cdots + b_k + a) - f(A + b_1 + \cdots + b_k) && (4.85)\\
&= f(B + a) - f(B) && (4.86)
\end{aligned}$$

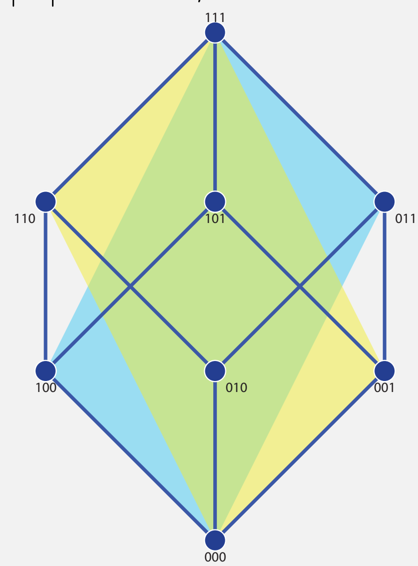
Submodular on Hypercube Vertices

- Test submodularity via values on verticies of hypercube.

Example: with $|V| = n = 2$, this is easy:



With $|V| = n = 3$, a bit harder.



How many inequalities?

Submodular Definitions

Theorem 4.6.3

Given function $f : 2^V \to \mathbb{R}$, then

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B) \text{ for all } A, B \subseteq V \qquad \text{(SC)}$$

if and only if

$$f(v|X) \geq f(v|Y) \text{ for all } X \subseteq Y \subseteq V \text{ and } v \notin Y \qquad \text{(DR)}$$

Proof.

(SC)⇒(DR): Set $A \leftarrow X \cup \{v\}$, $B \leftarrow Y$. Then $A \cup B = B \cup \{v\}$ and $A \cap B = X$ and $f(A) - f(A \cap B) \geq f(A \cup B) - f(B)$ implies (DR).

(DR)⇒(SC): Order $A \setminus B = \{v_1, v_2, \ldots, v_r\}$ arbitrarily. For $i \in 1 : r$,

$$f(v_i | (A \cap B) \cup \{v_1, v_2, \ldots, v_{i-1}\}) \geq f(v_i | B \cup \{v_1, v_2, \ldots, v_{i-1}\}).$$

Applying telescoping summation to both sides, we get:

$$\sum_{i=1}^{r} f(v_i | (A \cap B) \cup \{v_1, v_2, \ldots, v_{i-1}\}) \geq \sum_{i=1}^{r} f(v_i | B \cup \{v_1, v_2, \ldots, v_{i-1}\})$$

$$\Rightarrow \qquad f(A) - f(A \cap B) \geq f(A \cup B) - f(B)$$

Use of gain: submodular bounds of a difference

- Given submodular f, and given you have $C, D \subseteq E$ with either $D \supseteq C$ or $D \subseteq C$, and have an expression of the form:

$$f(C) - f(D) \tag{4.87}$$

- If $D \supseteq C$, then for any X with $D = C \cup X$ then

$$f(C) - f(D) = f(C) - f(C \cup X) \geq f(C \cap X) - f(X) \tag{4.88}$$

or

$$f(C \cup X|C) \leq f(X|C \cap X) \tag{4.89}$$

- Alternatively, if $D \subseteq C$, given any Y such that $D = C \cap Y$ then

$$f(C) - f(D) = f(C) - f(C \cap Y) \geq f(C \cup Y) - f(Y) \tag{4.90}$$

or

$$f(C|C \cap Y) \geq f(C \cup Y|Y) \tag{4.91}$$

- Equations (4.89) and (4.91) have same form.

Many (Equivalent) Definitions of Submodularity

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B), \quad \forall A, B \subseteq V \tag{4.92}$$

$$f(j|S) \geq f(j|T), \ \forall S \subseteq T \subseteq V, \text{ with } j \in V \setminus T \tag{4.93}$$

$$f(C|S) \geq f(C|T), \forall S \subseteq T \subseteq V, \text{ with } C \subseteq V \setminus T \tag{4.94}$$

$$f(j|S) \geq f(j|S \cup \{k\}), \ \forall S \subseteq V \text{ with } j \in V \setminus (S \cup \{k\}) \tag{4.95}$$

$$f(A \cup B|A \cap B) \leq f(A|A \cap B) + f(B|A \cap B), \quad \forall A, B \subseteq V \tag{4.96}$$

$$f(T) \leq f(S) + \sum_{j \in T \setminus S} f(j|S) - \sum_{j \in S \setminus T} f(j|S \cup T - \{j\}), \ \forall S, T \subseteq V \tag{4.97}$$

$$f(T) \leq f(S) + \sum_{j \in T \setminus S} f(j|S), \ \forall S \subseteq T \subseteq V \tag{4.98}$$

$$f(T) \leq f(S) - \sum_{j \in S \setminus T} f(j|S \setminus \{j\}) + \sum_{j \in T \setminus S} f(j|S \cap T) \ \forall S, T \subseteq V \tag{4.99}$$

$$f(T) \leq f(S) - \sum_{j \in S \setminus T} f(j|S \setminus \{j\}), \ \forall T \subseteq S \subseteq V \tag{4.100}$$

Equivalent Definitions of Submodularity

We've already seen that Eq. 4.92 $\equiv$ Eq. 4.93 $\equiv$ Eq. 4.94 $\equiv$ Eq. 4.95 $\equiv$ Eq. 4.96.
We next show that Eq. 4.95 $\Rightarrow$ Eq. 4.97 $\Rightarrow$ Eq. 4.98 $\Rightarrow$ Eq. 4.95.

Approach

To show these next results, we essentially first use:

$$f(S \cup T) = f(S) + f(T|S) \leq f(S) + \text{upper-bound} \tag{4.101}$$

and

$$f(T) + \text{lower-bound} \leq f(T) + f(S|T) = f(S \cup T) \tag{4.102}$$

leading to

$$f(T) + \text{lower-bound} \leq f(S) + \text{upper-bound} \tag{4.103}$$

or

$$f(T) \leq f(S) + \text{upper-bound} - \text{lower-bound} \tag{4.104}$$

Eq. 4.95 $\Rightarrow$ Eq. 4.97

Let $T \setminus S = \{j_1, \ldots, j_r\}$ and $S \setminus T = \{k_1, \ldots, k_q\}$.
First, we upper bound the gain of T in the context of S:

$$f(S \cup T) - f(S) = \sum_{t=1}^{r} \Big(f(S \cup \{j_1, \ldots, j_t\}) - f(S \cup \{j_1, \ldots, j_{t-1}\}) \Big) \tag{4.105}$$

$$= \sum_{t=1}^{r} f(j_t | S \cup \{j_1, \ldots, j_{t-1}\}) \leq \sum_{t=1}^{r} f(j_t | S) \tag{4.106}$$

$$= \sum_{j \in T \setminus S} f(j|S) \tag{4.107}$$

or

$$f(T|S) \leq \sum_{j \in T \setminus S} f(j|S) \tag{4.108}$$

Eq. 4.95 $\Rightarrow$ Eq. 4.97

Let $T \setminus S = \{j_1, \ldots, j_r\}$ and $S \setminus T = \{k_1, \ldots, k_q\}$.
Next, lower bound S in the context of T:

$$f(S \cup T) - f(T) = \sum_{t=1}^{q} [f(T \cup \{k_1, \ldots, k_t\}) - f(T \cup \{k_1, \ldots, k_{t-1}\})] \tag{4.109}$$

$$= \sum_{t=1}^{q} f(k_t | T \cup \{k_1, \ldots, k_t\} \setminus \{k_t\}) \geq \sum_{t=1}^{q} f(k_t | T \cup S \setminus \{k_t\}) \tag{4.110}$$

$$= \sum_{j \in S \setminus T} f(j | S \cup T \setminus \{j\}) \tag{4.111}$$

Eq. 4.95 $\Rightarrow$ Eq. 4.97

Let $T \setminus S = \{j_1, \dots, j_r\}$ and $S \setminus T = \{k_1, \dots, k_q\}$.
So we have the upper bound

$$f(T|S) = f(S \cup T) - f(S) \leq \sum_{j \in T \setminus S} f(j|S) \tag{4.112}$$

and the lower bound

$$f(S|T) = f(S \cup T) - f(T) \geq \sum_{j \in S \setminus T} f(j|S \cup T \setminus \{j\}) \tag{4.113}$$

This gives upper and lower bounds of the form

$$f(T) + \text{lower bound} \leq f(S \cup T) \leq f(S) + \text{upper bound}, \tag{4.114}$$

and combining directly the left and right hand side gives the desired inequality.

Eq. 4.97 $\Rightarrow$ Eq. 4.98

This follows immediately since if $S \subseteq T$, then $S \setminus T = \emptyset$, and the last term of Eq. 4.97 vanishes.

Eq. 4.98 $\Rightarrow$ Eq. 4.95

Here, we set $T = S \cup \{j, k\}$, $j \notin S \cup \{k\}$ into Eq. 4.98 to obtain

$$
\begin{aligned}
f(S \cup \{j,k\}) &\leq f(S) + f(j|S) + f(k|S) && (4.115)\\
&= f(S) + f(S + \{j\}) - f(S) + f(S + \{k\}) - f(S) && (4.116)\\
&= f(S + \{j\}) + f(S + \{k\}) - f(S) && (4.117)\\
&= f(j|S) + f(S + \{k\}) && (4.118)
\end{aligned}
$$

giving

$$
\begin{aligned}
f(j|S \cup \{k\}) &= f(S \cup \{j,k\}) - f(S \cup \{k\}) && (4.119)\\
&\leq f(j|S) && (4.120)
\end{aligned}
$$

Submodular Concave

- Why do we call the $f(A) + f(B) \geq f(A \cup B) + f(A \cap B)$ definition of submodularity, submodular concave?
- A continuous twice differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is concave iff $\nabla^2 f \preceq 0$ (the Hessian matrix is nonpositive definite).
- Define a "discrete derivative" or difference operator defined on discrete functions $f : 2^V \to \mathbb{R}$ as follows:

$$(\nabla_B f)(A) \triangleq f(A \cup B) - f(A \setminus B) = f\big(B|(A \setminus B)\big) \quad (4.121)$$

read as: the derivative of f at A in the direction B.

- Hence, if $A \cap B = \emptyset$, then $(\nabla_B f)(A) = f(B|A)$.
- Consider a form of second derivative or 2nd difference:

$$
\begin{aligned}
(\nabla_C \nabla_B f)(A) &= \nabla_C[\overbrace{f(A \cup B) - f(A \setminus B)}^{(\nabla_B f)(A)}] && (4.122)\\
&= (\nabla_B f)(A \cup C) - (\nabla_B f)(A \setminus C) && (4.123)\\
&= f(A \cup B \cup C) - f((A \cup C) \setminus B)\\
&\quad - f((A \setminus C) \cup B) + f((A \setminus C) \setminus B) && (4.124)
\end{aligned}
$$

Submodular Concave

- If the second difference operator everywhere nonpositive:

$$f(A \cup B \cup C) - f((A \cup C) \setminus B) - f((A \setminus C) \cup B) + f(A \setminus C \setminus B) \leq 0 \quad (4.125)$$

then we have the equation:

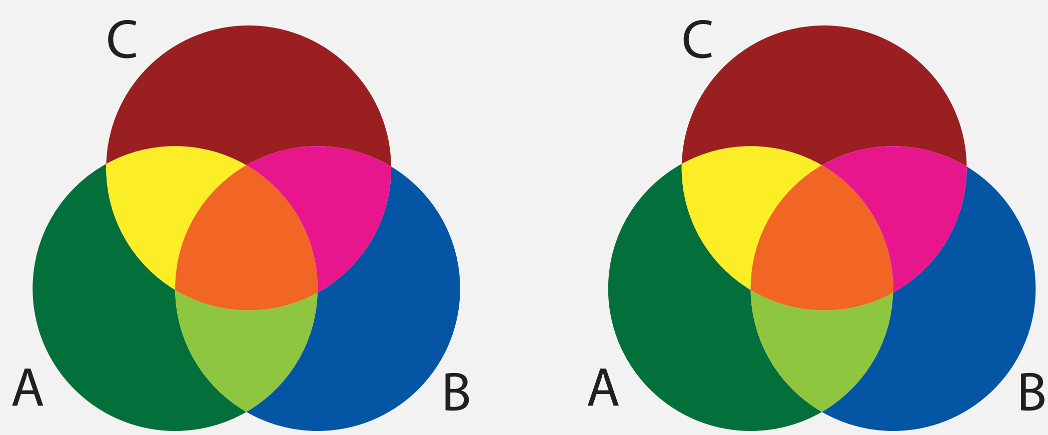
$$f((A \cup C) \setminus B) + f((A \setminus C) \cup B) \geq f(A \cup B \cup C) + f(A \setminus C \setminus B) \quad (4.126)$$

- Define $A' = (A \cup C) \setminus B$ and $B' = (A \setminus C) \cup B$. Then the above implies:

$$f(A') + f(B') \geq f(A' \cup B') + f(A' \cap B') \quad (4.127)$$

and note that A' and B' so defined can be arbitrary.
- One sense in which submodular functions are like concave functions.

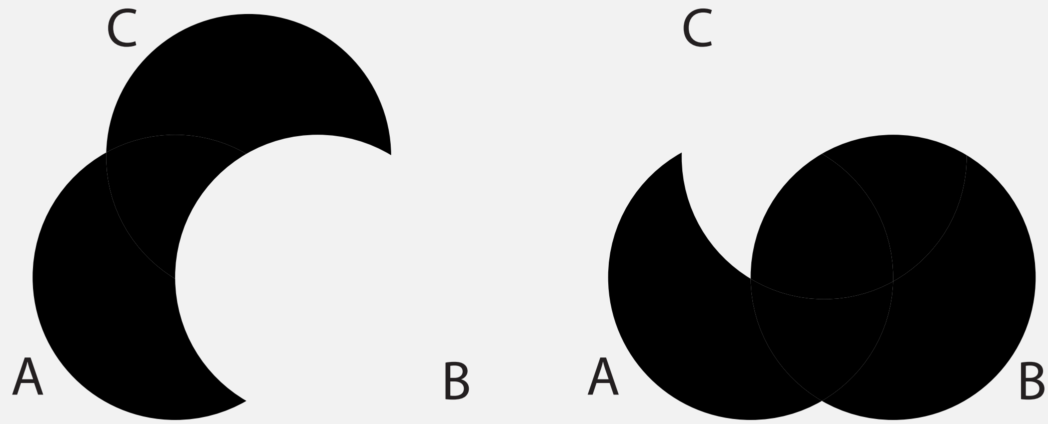
Submodular Concave



(a) $A' = (A \cup C) \setminus B$ (b) $B' = (A \setminus C) \cup B$

Figure: A figure showing $A' \cup B' = A \cup B \cup C$ and $A' \cap B' = A \setminus C \setminus B$.

Submodular Concave



(a) $A' = (A \cup C) \setminus B$

(b) $B' = (A \setminus C) \cup B$

Figure: A figure showing $A' \cup B' = A \cup B \cup C$ and $A' \cap B' = A \setminus C \setminus B$.

Submodularity and Concave

- This submodular/concave relationship is more simply done with singletons.
- Recall four points definition: A function is submodular if for all $X \subseteq V$ and $j, k \in V \setminus X$

$$f(X + j) + f(X + k) \geq f(X + j + k) + f(X) \tag{4.128}$$

- This gives us a simpler notion corresponding to concavity.
- Define gain as $\nabla_j(X) = f(X + j) - f(X)$, a form of discrete gradient.
- Trivially becomes a second-order condition, akin to concave functions: A function is submodular if for all $X \subseteq V$ and $j, k \in V$, we have:

$$\nabla_j \nabla_k f(X) \leq 0 \tag{4.129}$$