Submodular Functions, Optimization, and Applications to Machine Learning

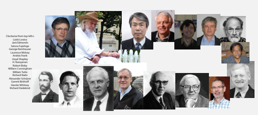
— Spring Quarter, Lecture 4 —

http://www.ee.washington.edu/people/faculty/bilmes/classes/ee563_spring_2018/

Prof. Jeff Bilmes

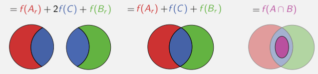
University of Washington, Seattle
Department of Electrical Engineering
http://melodi.ee.washington.edu/~bilmes

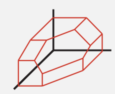
April 4th, 2018



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

$= f(A_r) + 2f(C) + f(B_r)$ $= f(A_r) + f(C) + f(B_r)$ $= f(A \cap B)$





Cumulative Outstanding Reading

- Read chapter 1 from Fujishige's book.

Announcements, Assignments, and Reminders

- Homework 1 out, due Monday, 4/9/2018 11:59pm electronically via our assignment dropbox (`https://canvas.uw.edu/courses/1216339/assignments`).
- If you have any questions about anything, please ask then via our discussion board (`https://canvas.uw.edu/courses/1216339/discussion_topics`).

Class Road Map - EE563

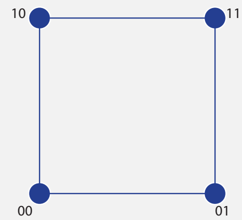
- L1(3/26): Motivation, Applications, & Basic Definitions,
- L2(3/28): Machine Learning Apps (diversity, complexity, parameter, learning target, surrogate).
- L3(4/2): Info theory exs, more apps, definitions, graph/combinatorial examples
- L4(4/4): Graph and Combinatorial Examples, Matrix Rank, Examples and Properties, visualizations
- L5(4/9):
- L6(4/11):
- L7(4/16):
- L8(4/18):
- L9(4/23):
- L10(4/25):
- L11(4/30):
- L12(5/2):
- L13(5/7):
- L14(5/9):
- L15(5/14):
- L16(5/16):
- L17(5/21):
- L18(5/23):
- L–(5/28): Memorial Day (holiday)
- L19(5/30):
- L21(6/4): Final Presentations maximization.

Last day of instruction, June 1st. Finals Week: June 2-8, 2018.

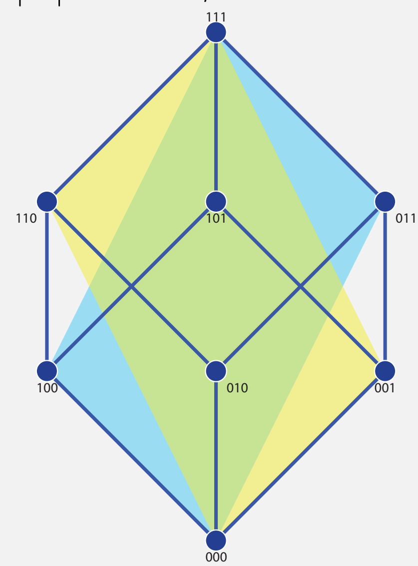
Submodular on Hypercube Vertices

- 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). □

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 $g : 2^V \to \mathbb{R}$ is supermodular, then so is $\bar{g} : 2^V \to \mathbb{R}$ defined as $\bar{g}(S) = g(V \setminus S)$.
- 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 number of connected components in G when we remove A; supermodular monotone non-decreasing but not normalized.

Graph Strength

- So $\bar{c}(A) = c(E \setminus A)$, the number of connected components in G when we remove A, is supermodular.
- Maximizing $\bar{c}(A)$ would be 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, $\textit{strength}(G, w)$.
- 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 $-w(E(X))$ 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:

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

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

□

Set Cover and Maximum Coverage
just Special cases of Submodular Optimization

- We are given a finite set U of m elements and a set of subsets $\mathcal{U} = \{U_1, U_2, \ldots, U_n\}$ of n subsets of U, so that $U_i \subseteq U$ and $\bigcup_i U_i = U$.
- The goal of minimum set cover is to choose the smallest subset $A \subseteq [n] \triangleq \{1, \ldots, n\}$ such that $\bigcup_{a \in A} U_a = U$.
- Maximum k cover: The goal in maximum coverage is, given an integer $k \leq n$, select k subsets, say $\{a_1, a_2, \ldots, a_k\}$ with $a_i \in [n]$ such that $|\bigcup_{i=1}^k U_{a_i}|$ is maximized.
- $f : 2^{[n]} \to \mathbb{Z}_+$ where for $A \subseteq [n]$, $f(A) = |\bigcup_{a \in A} U_a|$ is the set cover function and is submodular.
- Weighted set cover: $f(A) = w(\bigcup_{a \in A} U_a)$ where $w : U \to \mathbb{R}_+$.
- 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 $V \setminus X$ — i.e., $\delta(x) = E(X, V \setminus X)$.
- 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 $\{v, v_1, v_2, \dots, v_k\}$, 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 $x_{v_1}, \dots, x_{v_k}$.
- 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 $\mathbf{X} = (x_1, x_2, \dots, x_m)$ with $x_i \in \mathbb{R}^n$ for all i. There are m length-n column vectors $\{x_i\}_i$
- Let $V = \{1, 2, \dots, m\}$ 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

Example: Rank function of a matrix

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

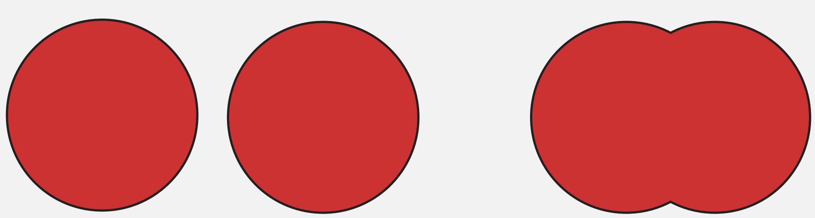
$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array}\begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{pmatrix} 0 & 2 & 2 & 3 & 0 & 1 & 3 & 1 \\ 0 & 3 & 0 & 4 & 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 & 3 & 0 & 0 & 5 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 5 \end{pmatrix} \end{array}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

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

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 all dimensions common to A and B. We call C the common span and call $A \cap B$ the common index.
- 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.,

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

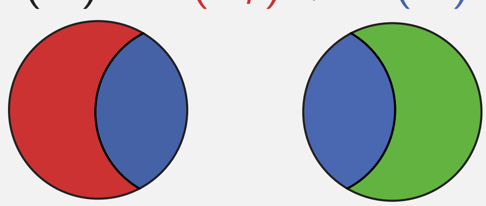
- But $r(A \cup B)$ counts the dimensions spanned by C only once.

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

Rank functions of a matrix

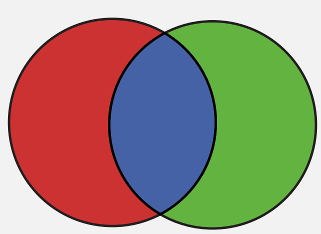
- Then $r(A) + r(B)$ counts the dimensions spanned by C twice, i.e.,

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



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

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

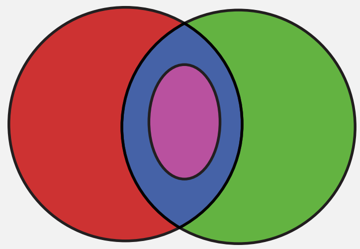


- Thus, we have subadditivity: $r(A) + r(B) \geq r(A \cup B)$. Can we add more to the r.h.s. and still have an inequality? Yes.

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).

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

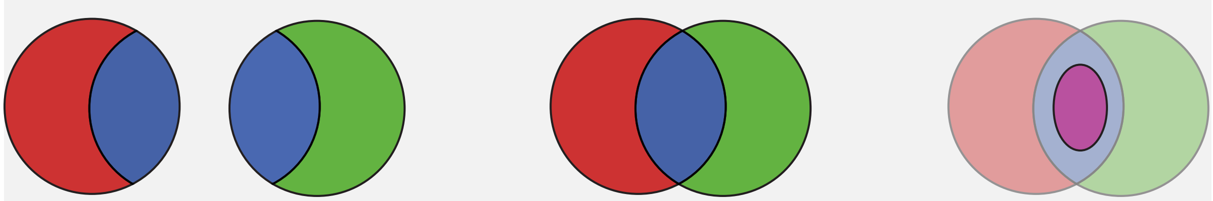


In short:

- 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

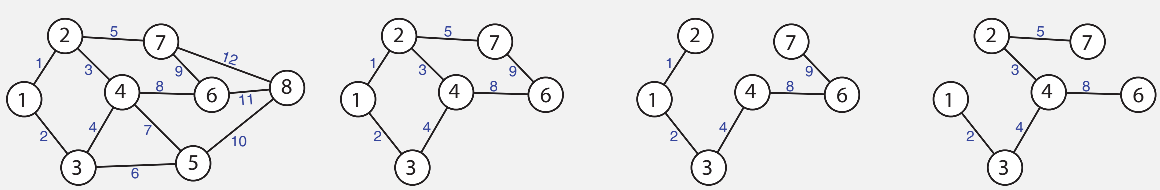
- Let S be a set of subspaces of a linear space (i.e., each $s \in S$ is a subspace of dimension ≥ 1).
- For each $X \subseteq S$, let $f(X)$ denote the dimensionality of the linear subspace spanned by the subspaces in X.
- We can think of S as a set of sets of vectors from the matrix rank example, and 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 $A \subseteq B$).
- We use the term non-decreasing rather than increasing, the latter of which is strict (also so that a constant function isn't "increasing").

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 $G = (V, E)$, $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$, $E = \{1, 2, \dots, 12\}$. $S = \{1, 2, 3, 4, 5, 8, 9\} \subset E$. 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.

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

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

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

- 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

$$f(A) + f(B) = f_1(A) + f_2(A) + f_1(B) + f_2(B) \tag{4.17}$$

$$\geq f_1(A \cup B) + f_2(A \cup B) + f_1(A \cap B) + f_2(A \cap B) \tag{4.18}$$

$$= f(A \cup B) + f(A \cap B). \tag{4.19}$$

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

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

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

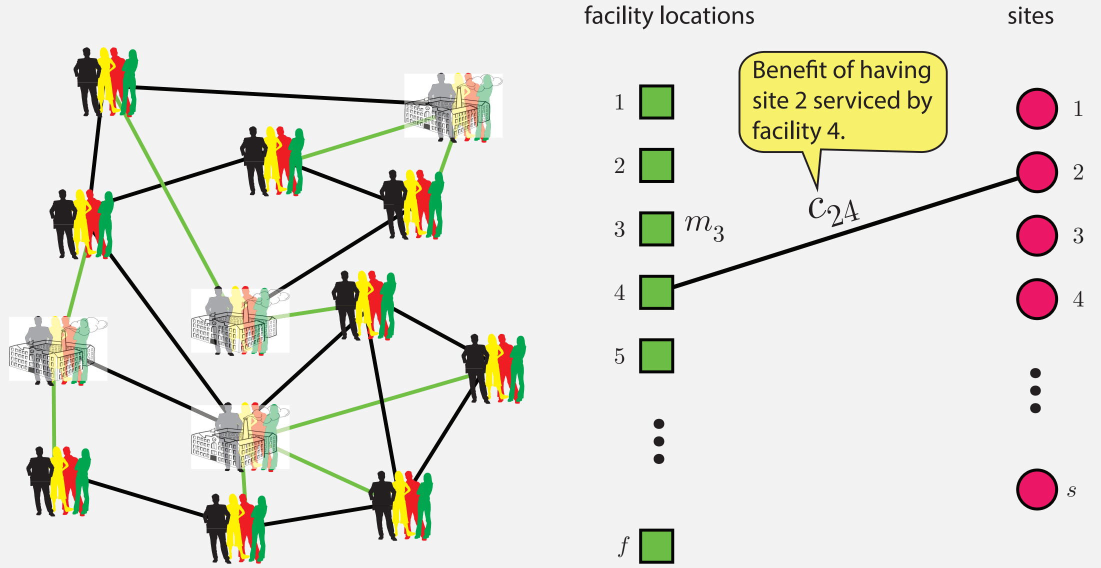
- Let $F = \{1, \ldots, f\}$ be a set of possible factory/plant locations for facilities to be built.
- $S = \{1, \ldots, s\}$ is a set of sites (e.g., cities, clients) needing service.
- Let c_{ij} be the "benefit" (e.g., $1/c_{ij}$ is the cost) of servicing site i with facility location j.
- Let m_j 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

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

- 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 w. $g(0) = 0$, 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 $f(A) \leq f(B)$ (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 $f(A) \geq f(B)$ (resp. $f(A) > f(B)$).

Composition of non-decreasing 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 non-decreasing or monotone non-increasing 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

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

or

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

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 non-decreasing Hence, $f(X \cup Y) + g(Y) - f(Y) \geq g(X \cup Y)$ giving

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

□

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

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

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

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

is submodular.

Proof.

For constant k, we have that $(f - k)$ is non-decreasing (or non-increasing) 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 $h_\alpha : 2^V \to \mathbb{R}$ as

$$h_\alpha(A) = \frac{1}{2}\Big(\min(\alpha, f(A)) + \min(\alpha, g(A))\Big) \tag{4.53}$$

  then h_α is submodular, and $h_\alpha(A) \geq \alpha$ 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 about it).

Arbitrary functions: difference between submodular funcs.

Theorem 4.5.5

Given an arbitrary set function h, it can be expressed as a difference between two submodular functions (i.e., $\forall h \in 2^V \to \mathbb{R}$, $\exists f, g$ s.t. $\forall A, h(A) = f(A) - g(A)$ where both f and g are submodular).

Proof.

Let h be given and arbitrary, and define:

$$\alpha \triangleq \min_{X,Y : X \not\subseteq Y, Y \not\subseteq X} \Big(h(X) + h(Y) - h(X \cup Y) - h(X \cap Y)\Big) \tag{4.54}$$

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

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

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

Arbitrary functions as difference between submodular funcs.

...cont.

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

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

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

□

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:

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

- 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

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

Inspired from information theory notation and the notation used for conditional entropy $H(X_A|X_B) = H(X_A, X_B) - H(X_B)$.

Totally normalized functions

- Any normalized submodular function g (even non-monotone) can be represented as a sum of a polymatroid (normalized monotone non-decreasing submodular) function $\bar{g}$ and a modular function m_g.
- Given arbitrary normalized submodular $g : 2^V \to \mathbb{R}$, construct a function $\bar{g} : 2^V \to \mathbb{R}$ as follows:

$$\bar{g}(A) = g(A) - \sum_{a \in A} g(a|V \setminus \{a\}) = g(A) - m_g(A) \tag{4.64}$$

  where $m_g(A) \triangleq \sum_{a \in A} g(a|V \setminus \{a\})$ is a modular function.
- $\bar{g}$ is normalized since $\bar{g}(\emptyset) = 0$.
- $\bar{g}$ is monotone non-decreasing since for $v \notin A \subseteq V$:

$$\bar{g}(v|A) = g(v|A) - g(v|V \setminus \{v\}) \geq 0 \tag{4.65}$$

- $\bar{g}$ is called the totally normalized version of g.
- Then $g(A) = \bar{g}(A) + m_g(A)$.

Arbitrary function as difference between two polymatroids

- Any normalized function h (i.e., $h(\emptyset) = 0$) can be represented as a difference not only between submodular, but between polymatroid (normalized monotone non-decreasing submodular) functions.
- Given submodular f and g, let $\bar{f}$ and $\bar{g}$ be them totally normalized.
- Given arbitrary $h = f - g$ where f and g are normalized submodular,

$$h = f - g = \bar{f} + m_f - (\bar{g} + m_g) \quad (4.66)$$

$$= \bar{f} - \bar{g} + (m_f - m_g) \quad (4.67)$$

$$= \bar{f} - \bar{g} + m_{f-h} \quad (4.68)$$

$$= \bar{f} + m^+_{f-g} - (\bar{g} + (-m_{f-g})^+) \quad (4.69)$$

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