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

\[ f(A) + f(B) \geq f(A \cup B) + f(A \cap B) \]
Read chapter 1 from Fujishige’s book.
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 discussion board (https://canvas.uw.edu/courses/1039754/discussion_topics)).
Class Road Map - IT-I

L1(3/28): Motivation, Applications, & Basic Definitions
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:

$$c_{ij} + c_{rs} \leq c_{is} + c_{rj}$$  \hspace{1cm} (4.15)

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

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

$$c_{\min(i,r),\min(j,s)} + c_{\max(i,r),\max(j,s)} \leq c_{is} + c_{rj}$$  \hspace{1cm} (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 vertices of hypercube.

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

```
00 01
11
```

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

```
000 001 010 011 100 101 110 111
```

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:

\[
 f(A) + f(B) \geq f(A \cup B) \quad (4.21)
\]

This means that the “whole” is less than the sum of the parts.
Superadditive Definitions

Definition 4.2.1 (superadditive)

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

$$f(A) + f(B) \leq f(A \cup B) \quad (4.21)$$

- 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 $f : 2^V \rightarrow \mathbb{R}_+$ where:

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

- 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, then for any \( A, B \subseteq V \), we have

\[
  f(A) + f(B) = f(A \cap B) + f(A \cup B)
\]

(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

\[
  f(A) = f(\emptyset) + \sum_{a \in A} \left( f(\{a\}) - f(\emptyset) \right) = c + \sum_{a \in A} f'(a)
\]

(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 $\overline{f} : 2^V \to \mathbb{R}$ as $\overline{f}(A) = f(V \setminus A)$ for any $A$.

**Proposition 4.2.1**

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

**Proof.**

\[
\overline{f}(A) + \overline{f}(B) \geq \overline{f}(A \cup B) + \overline{f}(A \cap B) \tag{4.26}
\]

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 \( f : 2^V \to \mathbb{R} \) 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.,

\[
    c(A + a) - c(A) \leq c(B + a) - c(B)
\]

(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 connected 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 \( \overline{c}(A) = c(E \setminus A) \) is the number of connected components in \( G \) when we remove \( A \), is supermodular.
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).
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.

\[
\text{Weak} \equiv \exists A \text{ with } |A| \text{ small } \bar{c}(A) \text{ big.}
\]
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.
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 $w : E \rightarrow \mathbb{R}^+$ be a weighted graph with non-negative weights.
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 $w : E \rightarrow \mathbb{R}^+$ 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

\[
    w(A) = \sum_{e \in A} w_e
\]  

(4.1)

so that \( w(E(G[S])) \) is the “internal strength” of the vertex set \( S \).

Notation: \( S \) is a set of nodes, \( G[S] \) is the vertex-induced subgraph of \( G \) induced by vertices \( S \), \( E(G[S]) \) are the edges contained within this induced subgraph, and \( w(E(G[S])) \) is the weight of these edges.
Graph Strength

- Then $w(A)$ for $A \subseteq E$ is a modular function
  
  $$w(A) = \sum_{e \in A} w_e$$  \hspace{1cm} (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 $\overline{c}(A) > 1$ components —
Graph Strength

- Then $w(A)$ for $A \subseteq E$ is a modular function
  $$w(A) = \sum_{e \in A} w_e$$  \hspace{1cm} (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 $\overline{c}(A) > 1$ components — then $w(A)/(\overline{c}(A) - 1)$ is like the “effort per achieved/additional component” for a network attacker.
Graph Strength

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

\[
    w(A) = \sum_{e \in A} w_e
\]

(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:

\[
    \text{strength}(G, w) = \min_{A \subseteq E(G): \bar{c}(A) > 1} \frac{w(A)}{\bar{c}(A) - 1}
\]

(4.2)
Graph Strength

- Then $w(A)$ for $A \subseteq E$ is a modular function
  
  $$w(A) = \sum_{e \in A} w_e$$  
  
  (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:

  $$\text{strength}(G, w) = \min_{A \subseteq E(G) : \bar{c}(A) > 1} \frac{w(A)}{\bar{c}(A) - 1}$$  
  
  (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, $\text{strength}(G, w)$. 
Graph Strength

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

\[
w(A) = \sum_{e \in A} w_e
\]  \hspace{1cm} (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:

\[
\text{strength}(G, w) = \min_{A \subseteq E(G) : \bar{c}(A) > 1} \frac{w(A)}{\bar{c}(A) - 1}
\]  \hspace{1cm} (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, \( \text{strength}(G, w) \).

- Since submodularity, problems have strongly-poly-time solutions.
Lemma 4.3.1

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

$$f(X) = m^T 1_X + \frac{1}{2} 1_X^T M 1_X$$

(4.3)

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

Proof.
Lemma 4.3.1

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

\[
f(X) = \mathbf{m}^\top \mathbf{1}_X + \frac{1}{2} \mathbf{1}_X^\top \mathbf{M} \mathbf{1}_X
\]

is submodular iff the off-diagonal elements of \( \mathbf{M} \) 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.

\[
|E(X)| = \sum_{x \in X} 1
\]

where \( x \in g \in X \)
Lemma 4.3.1

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

$$f(X) = m^T 1_X + \frac{1}{2} 1_X^T M 1_X$$

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.

$$w(E(X)) = \sum_{x, y \in X} w(x, y)$$
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

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

(4.3)

is submodular iff the off-diagonal elements of $\mathbf{M}$ 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 $m^\top \mathbf{1}_A = m(A)$ added to a weighted submodular function, hence $f$ is submodular.
Proof of Lemma 4.3.1 cont.

- Conversely, suppose $f$ is submodular.
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 $f(\emptyset) = 0$. 
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:

$$0 \leq f(\{u\}) + f(\{v\}) - f(\{u, v\}) = m(u) + \frac{1}{2} M_{u,u} + m(v) + \frac{1}{2} M_{v,v} - \left( m(u) + m(v) + \frac{1}{2} M_{u,u} + M_{u,v} + \frac{1}{2} M_{v,v} \right)$$

$$= -M_{u,v}$$

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

□
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 $\bigcup_i V_i = V$.\[\]
**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 $\bigcup_{a \in A} V_a = V$. 

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 $k \leq m$, 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.
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 $k \leq m$, 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.
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 \( V = \{V_1, V_2, \ldots, V_m\} \) 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 \( k \leq m \), 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.

- \( f : 2^[[m]] \rightarrow \mathbb{Z}_+ \) 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.
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 \( V \setminus S \).
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 $V \setminus S$. 
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 $\delta(x) = E(X, V \setminus X)$. 


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 $\delta : 2^V \rightarrow \mathbb{R}_+$ 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, $f(X) = w(\delta(X))$. 
**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 $m$).
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:

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

(4.8)

If not, then $x_v$ is linearly independent of $x_{v_1}, \ldots, x_{v_k}$.
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:

$$x_v = \sum_{i=1}^k \alpha_i x_{v_i} \quad (4.8)$$

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 $X = (x_1, x_2, \ldots, x_m)$ with $x_i \in \mathbb{R}^n$ for all $i$. There are $m$ length-$n$ column vectors $\{x_i\}_i$. 

- Skip matrix rank example
Example: Rank function of a matrix

- Given $n \times m$ matrix $X = (x_1, x_2, \ldots, x_m)$ 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.
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 $\{x_i\}$.
- 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 $A$. 
Example: Rank function of a matrix

- Given $n \times m$ matrix $X = (x_1, x_2, \ldots, x_m)$ 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 $\{x_a\}_{a \in A}$.
Given \( n \times m \) matrix \( X = (x_1, x_2, \ldots, x_m) \) with \( x_i \in \mathbb{R}^n \) for all \( i \).

There are \( m \) length-\( n \) column vectors \( \{x_i\} \).

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 \( X \).
Example: Rank function of a matrix

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

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

- Let $A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$, $C = \{6, 7\}$, $A_r = \{1\}$, $B_r = \{5\}$.
- Then $r(A) = 3$, $r(B) = 3$, $r(C) = 2$.
- $r(A \cup C) = 3$, $r(B \cup C) = 3$.
- $r(A \cup A_r) = 3$, $r(B \cup B_r) = 3$, $r(A \cup B_r) = 4$, $r(B \cup A_r) = 4$.
- $r(A \cup B) = 4$, $r(A \cap B) = 1 < r(C) = 2$. 
Example: Rank function of a matrix

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

Let $A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$, $C = \{6, 7\}$, $A_r = \{1\}$, $B_r = \{5\}$.

Then $r(A) = 3$, $r(B) = 3$, $r(C') = 2$.

$r(A \cup C') = 3$, $r(B \cup C') = 3$.

$r(A \cup A_r) = 3$, $r(B \cup B_r) = 3$, $r(A \cup B_r) = 4$, $r(B \cup A_r) = 4$.

$r(A \cup B) = 4$, $r(A \cap B) = 1 < r(C') = 2$. 
Example: Rank function of a matrix

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

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

- Let $A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$, $C = \{6, 7\}$, $A_r = \{1\}$, $B_r = \{5\}$.
- Then $r(A) = 3$, $r(B) = 3$, $r(C) = 2$.
- $r(A \cup C') = 3$, $r(B \cup C') = 3$.
- $r(A \cup A_r) = 3$, $r(B \cup B_r) = 3$, $r(A \cup B_r) = 4$, $r(B \cup A_r) = 4$.
- $r(A \cup B) = 4$, $r(A \cap B) = 1 < r(C) = 2$. 
Example: Rank function of a matrix

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

Let $A = \{1, 2, 3\}, \ B = \{3, 4, 5\}, \ C = \{6, 7\}, \ A_r = \{1\}, \ B_r = \{5\}.$

Then $r(A) = 3, \ r(B) = 3, \ r(C) = 2.$

$r(A \cup C') = 3, \ r(B \cup C') = 3.$

$r(A \cup A_r) = 3, \ r(B \cup B_r) = 3, \ r(A \cup B_r) = 4, \ r(B \cup A_r) = 4.$

$r(A \cup B) = 4, \ r(A \cap B) = 1 < r(C) = 2.$
Example: Rank function of a matrix

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

\[
\begin{pmatrix}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
1 & 2 & 2 & 3 & 0 & 1 & 3 & 1 \\
2 & 0 & 3 & 0 & 4 & 0 & 0 & 2 & 4 \\
3 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 5 \\
4 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 5
\end{pmatrix}
= \begin{pmatrix}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
1 & | & | & | & | & | & | & | \\
\times_1 & \times_2 & \times_3 & \times_4 & \times_5 & \times_6 & \times_7 & \times_8
\end{pmatrix}
\]

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

---

Prof. Jeff Bilmes
EE596b/Spring 2016/Submodularity - Lecture 4 - Apr 6th, 2016
F22/77 (pg.55/221)
Example: Rank function of a matrix

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

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

Let $A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$, $C = \{6, 7\}$, $A_r = \{1\}$, $B_r = \{5\}$.

Then $r(A) = 3$, $r(B) = 3$, $r(C') = 2$.

$r(A \cup C') = 3$, $r(B \cup C') = 3$.

$r(A \cup A_r) = 3$, $r(B \cup B_r) = 3$, $r(A \cup B_r) = 4$, $r(B \cup A_r) = 4$.

$r(A \cup B) = 4$, $r(A \cap B) = 1 < r(C) = 2$. 
**Example: Rank function of a matrix**

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

\[
\begin{pmatrix}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
1 & 0 & 2 & 2 & 3 & 0 & 1 & 3 & 1 \\
2 & 0 & 3 & 0 & 4 & 0 & 0 & 2 & 4 \\
3 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 5 \\
4 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 5 \\
\end{pmatrix}
= \begin{pmatrix}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
\mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid \\
x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 \\
\mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid \\
\end{pmatrix}
\]

- Let $A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$, $C = \{6, 7\}$, $A_r = \{1\}$, $B_r = \{5\}$.
- Then $r(A) = 3$, $r(B) = 3$, $r(C) = 2$.
- $r(A \cup C) = 3$, $r(B \cup C) = 3$.
- $r(A \cup A_r) = 3$, $r(B \cup B_r) = 3$, $r(A \cup B_r) = 4$, $r(B \cup A_r) = 4$.
- $r(A \cup B) = 4$, $r(A \cap B) = 1 < r(C) = 2$. 
Example: Rank function of a matrix

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

\[
\begin{pmatrix}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
0 & 2 & 2 & 3 & 0 & 1 & 3 & 1 \\
0 & 3 & 0 & 4 & 0 & 0 & 2 & 4 \\
0 & 0 & 0 & 0 & 3 & 0 & 0 & 5 \\
2 & 0 & 0 & 0 & 0 & 0 & 0 & 5 \\
\end{pmatrix}
= 
\begin{pmatrix}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
\mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid \\
\mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid \\
X_1 & X_2 & X_3 & X_4 & X_5 & X_6 & X_7 & X_8 \\
\mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid \\
\end{pmatrix}
\]

- Let $A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$, $C = \{6, 7\}$, $A_r = \{1\}$, $B_r = \{5\}$.
- Then $r(A) = 3$, $r(B) = 3$, $r(C) = 2$.
- $r(A \cup C) = 3$, $r(B \cup C) = 3$.
- $r(A \cup A_r) = 3$, $r(B \cup B_r) = 3$, $r(A \cup B_r) = 4$, $r(B \cup A_r) = 4$.
- $r(A \cup B) = 4$, $r(A \cap B) = 1 < r(C) = 2$. 
Example: Rank function of a matrix

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

\[
\begin{pmatrix}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
1 & 0 & 2 & 2 & 3 & 0 & 1 & 3 & 1 \\
2 & 0 & 3 & 0 & 4 & 0 & 0 & 2 & 4 \\
3 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 5 \\
4 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 5 \\
\end{pmatrix}
= 
\begin{pmatrix}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
\mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid \\
x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 \\
\mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid \\
\end{pmatrix}
\]

- Let $A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$, $C = \{6, 7\}$, $A_r = \{1\}$, $B_r = \{5\}$.
- Then $r(A) = 3$, $r(B) = 3$, $r(C) = 2$.
- $r(A \cup C') = 3$, $r(B \cup C') = 3$.
- $r(A \cup A_r) = 3$, $r(B \cup B_r) = 3$, $r(A \cup B_r) = 4$, $r(B \cup A_r) = 4$.
- $r(A \cup B) = 4$, $r(A \cap B) = 1 < r(C) = 2$. 
**Example: Rank function of a matrix**

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

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

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

Let $A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$, $C = \{6, 7\}$, $A_r = \{1\}$, $B_r = \{5\}$.

Then $r(A) = 3$, $r(B) = 3$, $r(C) = 2$.

$r(A \cup C) = 3$, $r(B \cup C) = 3$.

$r(A \cup A_r) = 3$, $r(B \cup B_r) = 3$, $r(A \cup B_r) = 4$, $r(B \cup A_r) = 4$.

$r(A \cup B) = 4$, $r(A \cap B) = 1 < r(C) = 2$. 

---

*Prof. Jeff Bilmes*

EE596b/Spring 2016/Submodularity - Lecture 4 - Apr 6th, 2016

F22/77 (pg.60/221)
Example: Rank function of a matrix

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

\[
\begin{pmatrix}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
1 & 0 & 2 & 3 & 0 & 1 & 3 & 1 \\
2 & 0 & 3 & 0 & 4 & 0 & 0 & 2 \\
3 & 0 & 0 & 0 & 0 & 3 & 0 & 0 \\
4 & 2 & 0 & 0 & 0 & 0 & 0 & 5 \\
\end{pmatrix} = \begin{pmatrix}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
\mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid \\
x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 \\
\mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid \\
\end{pmatrix}
\]

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

Prof. Jeff Bilmes
Example: Rank function of a matrix

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

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

- Let $A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$, $C = \{6, 7\}$, $A_r = \{1\}$, $B_r = \{5\}$.
- Then $r(A) = 3$, $r(B) = 3$, $r(C) = 2$.
- $r(A \cup C) = 3$, $r(B \cup C) = 3$.
- $r(A \cup A_r) = 3$, $r(B \cup B_r) = 3$, $r(A \cup B_r) = 4$, $r(B \cup A_r) = 4$.
- $r(A \cup B) = 4$, $r(A \cap B) = 1 < r(C) = 2$. 
**Example: Rank function of a matrix**

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

\[ \begin{pmatrix}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
0 & 2 & 2 & 3 & 0 & 1 & 3 & 1 \\
0 & 3 & 0 & 4 & 0 & 0 & 2 & 4 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 5 \\
2 & 0 & 0 & 0 & 0 & 0 & 0 & 5 \\
\end{pmatrix} = \begin{pmatrix}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 \\
\end{pmatrix} \]

- Let $A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$, $C = \{6, 7\}$, $A_r = \{1\}$, $B_r = \{5\}$.
- Then $\text{r}(A) = 3$, $\text{r}(B) = 3$, $\text{r}(C) = 2$.
- $\text{r}(A \cup C) = 3$, $\text{r}(B \cup C) = 3$.
- $\text{r}(A \cup A_r) = 3$, $\text{r}(B \cup B_r) = 3$, $\text{r}(A \cup B_r) = 4$, $\text{r}(B \cup A_r) = 4$.
- $\text{r}(A \cup B) = 4$, $\text{r}(A \cap B) = 1 < \text{r}(C) = 2$. 
Example: Rank function of a matrix

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

\[
\begin{pmatrix}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
0 & 2 & 2 & 3 & 0 & 1 & 3 & 1 \\
0 & 3 & 0 & 4 & 0 & 0 & 2 & 4 \\
0 & 0 & 0 & 0 & 3 & 0 & 0 & 5 \\
2 & 0 & 0 & 0 & 0 & 0 & 0 & 5 \\
\end{pmatrix}
= 
\begin{pmatrix}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
& & & & & & & \\
& & & & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 \\
& & & & & & & \\
& & & & & & & \\
& & & & & & & \\
& & & & & & & \\
\end{pmatrix}
\]

- Let $A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$, $C = \{6, 7\}$, $A_r = \{1\}$, $B_r = \{5\}$.
- Then $r(A) = 3$, $r(B) = 3$, $r(C) = 2$.
- $r(A \cup C) = 3$, $r(B \cup C) = 3$.
- $r(A \cup A_r) = 3$, $r(B \cup B_r) = 3$, $r(A \cup B_r) = 4$, $r(B \cup A_r) = 4$.
- $r(A \cup B) = 4$, $r(A \cap B) = 1 < r(C) = 2$. 
Example: Rank function of a matrix

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

$$
\begin{pmatrix}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
1 & 0 & 2 & 2 & 3 & 0 & 1 & 3 & 1 \\
2 & 0 & 3 & 0 & 4 & 0 & 0 & 2 & 4 \\
3 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 5 \\
4 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 5 \\
\end{pmatrix} = 
\begin{pmatrix}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
\vert & \vert & \vert & \vert & \vert & \vert & \vert & \vert & \vert \\
X_1 & X_2 & X_3 & X_4 & X_5 & X_6 & X_7 & X_8 \\
\vert & \vert & \vert & \vert & \vert & \vert & \vert & \vert & \vert \\
\end{pmatrix}
$$

- Let $A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$, $C = \{6, 7\}$, $A_r = \{1\}$, $B_r = \{5\}$.
- Then $r(A) = 3$, $r(B) = 3$, $r(C) = 2$.
- $r(A \cup C) = 3$, $r(B \cup C) = 3$.
- $r(A \cup A_r) = 3$, $r(B \cup B_r) = 3$, $r(A \cup B_r) = 4$, $r(B \cup A_r) = 4$.
- $r(A \cup B) = 4$, $r(A \cap B) = 1 < r(C) = 2$. 
Example: Rank function of a matrix

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

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

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

Prof. Jeff Bilmes  
EE596b/Spring 2016/Submodularity - Lecture 4 - Apr 6th, 2016  
F22/77 (pg.66/221)
Example: Rank function of a matrix

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

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

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

Common index.
Example: Rank function of a matrix

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

\[
\begin{pmatrix}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
0 & 2 & 2 & 3 & 0 & 1 & 3 & 1 \\
0 & 3 & 0 & 4 & 0 & 0 & 2 & 4 \\
0 & 0 & 0 & 0 & 3 & 0 & 0 & 5 \\
2 & 0 & 0 & 0 & 0 & 0 & 0 & 5 \\
\end{pmatrix}
\begin{pmatrix}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 \\
\end{pmatrix}
\]

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

[common span.]

Prof. Jeff Bilmes
EE596b/Spring 2016/Submodularity - Lecture 4 - Apr 6th, 2016
F22/77 (pg.68/221)
Example: Rank function of a matrix

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

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

- 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.
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.
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.
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.

$$r(A) + r(B) \geq r(A \cup B)$$
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.

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

- 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.
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.

\[
 r(A) + r(B) \geq r(A \cup B) + \square
\]

- 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.
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 $B$. 

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.,

$r(A) + r(B) = r(A_r) + 2r(C) + r(B_r)$ (4.9)

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

$r(A \mid B) = r(A_r) + r(C) + r(B_r)$ (4.10)
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 $B$.

\[
\begin{align*}
\text{Let } r(A) &= r(C) + r(A_r) \\
\text{Similarly, } r(B) &= r(C) + r(B_r) \\
\text{Then, } r(A) + r(B) &= r(A_r) + 2r(C) + r(B_r) \quad (4.9) \\
\text{But } r(A\{B\}) &= r(A_r) + r(C) + r(B_r) \quad (4.10)
\end{align*}
\]
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 $A$. 
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, $r(A) = r(C) + r(A_r)$
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, $r(B) = r(C) + r(B_r)$. 

(4.9)

But $r(A[B])$ counts the dimensions spanned by $C$ only once.

$r(A[B]) = r(A_r) + r(C) + r(B_r)$  (4.10)
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.,

$$r(A) + r(B) = r(A_r) + 2r(C) + r(B_r).$$  \hspace{1cm} (4.9)
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.,

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

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

$$r(A \cup B) = r(A_r) + r(C) + r(B_r).$$  \(4.10\)
Then $r(A) + r(B)$ counts the dimensions spanned by $C$ twice, i.e.,

$$r(A) + r(B) = r(A_r) + 2r(C) + r(B_r)$$
Rank functions of a matrix

- Then $r(A) + r(B)$ counts the dimensions spanned by $C$ twice, i.e.,
  $$r(A) + r(B) = r(Ar) + 2r(C) + r(Br)$$

- But $r(A \cup B)$ counts the dimensions spanned by $C$ only once.
  $$r(A \cup B) = r(Ar) + r(C) + r(Br)$$
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:
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).
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

\[ r(A \cup B) \leq r(C) \]

\[
\begin{align*}
r(A) + r(B) & \geq r(A \cup B) + r(A \cap B) \\
&= r(A_r) + 2r(C) + r(B_r) \\
&= r(A_r) + r(C) + r(B_r) \\
&= r(A \cap B)
\end{align*}
\]
Let $S$ be a set of subspaces of a linear space (i.e., each $s \in S$ is a subspace of dimension $\geq 1$).
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 $\geq 1$).

- For each $X \subseteq S$, let $f(X)$ denote the dimensionality of the linear subspace spanned by the subspaces in $X$. 
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 $\geq 1$).
- 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.
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 $\geq 1$).
- 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 $f : 2^S \rightarrow \mathbb{R}_+$ as follows,
  
  $$f(X) = r(\bigcup_{s \in S} X_s)$$  
  
  (4.11)

we have that $f$ is submodular, and is known to be a 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 $\geq 1$).

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 $f : 2^S \rightarrow \mathbb{R}_+$ as follows,

$$f(X) = r(\bigcup_{s \in S} X_s)$$  \hspace{1cm} (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 $S$. 
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 $S$).
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 $G = (V, E)$, $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$, $E = \{1, 2, \ldots, 12\}$. $S = \{1, 2, 3, 4, 5, 8, 9\} \subseteq E$. 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.
Given $E$, let $f_1, f_2 : 2^E \rightarrow \mathbb{R}$ be two submodular functions. Then

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

(4.16)

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

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

is submodular. This follows easily since

$$f(A) + f(B) = f_1(A) + f_2(A) + f_1(B) + f_2(B)$$   (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)

I.e., it holds for each component of $f$ in each term in the inequality.
Given $E$, let $f_1, f_2 : 2^E \rightarrow \mathbb{R}$ be two submodular functions. Then

$$f : 2^E \rightarrow \mathbb{R} \text{ with } f(A) = f_1(A) + f_2(A)$$ \hspace{1cm} (4.16)

is submodular. This follows easily since

$$f(A) + f(B) = f_1(A) + f_2(A) + f_1(B) + f_2(B) \geq f_1(A \cup B) + f_2(A \cup B) + f_1(A \cap B) + f_2(A \cap B) \hspace{1cm} (4.17)$$

$$= f(A \cup B) + f(A \cap B). \hspace{1cm} (4.18)$$

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 $\alpha_1, \alpha_2 \geq 0$. 

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

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

(4.20)

is submodular (as is $f(A) = f_1(A) + m(A)$).
Given $E$, let $f_1, m : 2^E \to \mathbb{R}$ be a submodular and a modular function. Then

$$f : 2^E \to \mathbb{R} \text{ with } f(A) = f_1(A) - m(A)$$ \hspace{1cm} (4.20)

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

$$f(A) + f(B) = f_1(A) - m(A) + f_1(B) - m(B)$$ \hspace{1cm} (4.21)

$$\geq f_1(A \cup B) - m(A \cup B) + f_1(A \cap B) - m(A \cap B)$$ \hspace{1cm} (4.22)

$$= f(A \cup B) + f(A \cap B).$$ \hspace{1cm} (4.23)
Given $E$, let $f_1, m : 2^E \rightarrow \mathbb{R}$ be a submodular and a modular function. Then

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

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

$$f(A) + f(B) = f_1(A) - m(A) + f_1(B) - m(B) \quad (4.21)$$

$$\geq f_1(A \cup B) - m(A \cup B) + f_1(A \cap B) - m(A \cap B) \quad (4.22)$$

$$= f(A \cup B) + f(A \cap B). \quad (4.23)$$

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

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

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

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

(4.24)

is submodular.

**Proof.**

[Proof content]
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

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

is submodular.

Proof.

Given \( A \subseteq B \subseteq E \setminus v \), consider

\[
f((A + v) \cap S) - f(A \cap S) \geq f((B + v) \cap S) - f(B \cap S)
\]

(4.25)
Graph & Combinatorial Examples

Matrix Rank

Examples and Properties

Other SubmodularDefs.

Restricting Submodular Functions

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

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

(4.24)

is submodular.

Proof.

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

$$f((A + v) \cap S) - f(A \cap S) \geq f((B + v) \cap S) - f(B \cap S)$$

(4.25)

If $v \not\in S$, then both differences on each size are zero.
Restricting Submodular Functions

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

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

is submodular.

**Proof.**

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

$$f((A + v) \cap S) - f(A \cap S) \geq f((B + v) \cap S) - f(B \cap S) \quad (4.25)$$

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

$$\left( A + v \right) \cap S = (A \cap S) + v$$

$$f(A' + v) - f(A') \geq f(B' + v) - f(B') \quad (4.26)$$

with $A' = A \cap S$ and $B' = B \cap S$. Since $A' \subseteq B'$, this holds due to submodularity of $f$. \qed
Given $V$, let $f_1, f_2 : 2^V \rightarrow \mathbb{R}$ be two submodular functions and let $S_1, S_2$ be two arbitrary fixed sets. Then

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

(4.27)

is submodular. This follows easily from the preceding two results.
Summing Restricted Submodular Functions

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

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

is submodular. This follows easily from the preceding two results.

Given $V$, let $C = \{C_1, C_2, \ldots, C_k\}$ be a set of subsets of $V$, and for each $C \in C$, let $f_C : 2^V \rightarrow \mathbb{R}$ be a submodular function. Then

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

is submodular.
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

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

(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

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

(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.
Given $V$, let $c \in \mathbb{R}^V$ be a given fixed vector. Then $f : 2^V \rightarrow \mathbb{R}_+$, where
\[
f(A) = \max_{j \in A} c_j
\] (4.29)
is submodular and normalized (we take $f(\emptyset) = 0$).

Proof.

Consider
\[
\max_{j \in A} c_j + \max_{j \in B} c_j \geq \max_{j \in A \cup B} c_j + \max_{j \in A \cap B} c_j
\] (4.30)
which follows since we have that
\[
\max(\max_{j \in A} c_j, \max_{j \in B} c_j) = \max_{j \in A \cup B} c_j
\] (4.31)
and
\[
\min(\max_{j \in A} c_j, \max_{j \in B} c_j) \geq \max_{j \in A \cap B} c_j
\] (4.32)
Given $V$, let $c \in \mathbb{R}^V$ be a given fixed vector (not necessarily non-negative).

Then $f : 2^V \rightarrow \mathbb{R}$, where

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

(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.

<table>
<thead>
<tr>
<th>Facility/Plant Location (uncapacitated) w. plant benefits</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Let $F = {1, \ldots, f}$ be a set of possible factory/plant locations for facilities to be built.</td>
</tr>
</tbody>
</table>
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.
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 $F = \{1, \ldots, f\}$ 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 $j$. 
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.
• 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.
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

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

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

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/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/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.

We can model this with a weighted bipartite graph \( G = (F, S, E, c) \) where \( F \) is set of possible factory/plant locations, \( S \) is set of sites needing service, \( E \) are edges indicating (factory,site) service possibility pairs, and \( c : E \rightarrow \mathbb{R}_+ \) is the benefit of a given pair.

Facility location function has form:

\[
f(A) = \sum_{i \in F} \max_{j \in A} c_{ij}. \quad (4.35)
\]
Given $V, E$, let $c \in \mathbb{R}^{V \times E}$ be a given $|V| \times |E|$ matrix. Then

$$f : 2^E \to \mathbb{R}, \text{ where } f(A) = \sum_{i \in V} \max_{j \in A} c_{ij} \quad (4.36)$$

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 $\Sigma$ 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 $\Sigma_A$ be the (square) submatrix of $\Sigma$ obtained by including only entries in the rows/columns given by $A$. 

\[
\log \det (\Sigma_A) = f(A)
\] (4.37)
Log Determinant

- Let $\Sigma$ 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 $\Sigma_A$ be the (square) submatrix of $\Sigma$ obtained by including only entries in the rows/columns given by $A$.
- We have that:

$$f(A) = \log \det(\Sigma_A) \text{ is submodular.} \quad (4.37)$$
Log Determinant

- Let $\Sigma$ 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 $\Sigma_A$ be the (square) submatrix of $\Sigma$ obtained by including only entries in the rows/columns given by $A$.
- We have that:

$$f(A) = \log \det(\Sigma_A) \text{ is submodular.} \quad (4.37)$$

- The submodularity of the log determinant is crucial for determinantal point processes (DPPs) (defined later in the class).
Let $\Sigma$ 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 $\Sigma_A$ be the (square) submatrix of $\Sigma$ obtained by including only entries in the rows/columns given by $A$.

We have that:

$$f(A) = \log \det (\Sigma_A) \text{ is submodular.} \quad (4.37)$$

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

$$x \in p(x) = \frac{1}{\sqrt{|2\pi \Sigma|}} \exp \left( -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right) \quad (4.38)$$
Then the (differential) entropy of the r.v. $X$ is given by

$$h(X) = \log \sqrt{|2\pi e\Sigma|} = \log \sqrt{(2\pi e)^n|\Sigma|} \quad (4.39)$$

and in particular, for a variable subset $A$,

$$f(A) = h(X_A) = \log \sqrt{(2\pi e)|A||\Sigma_A|} \quad (4.40)$$

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

$$f(A) = h(X_A) = m(A) + \frac{1}{2} \log |\Sigma_A| \quad (4.41)$$

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 $f_i : 2^V \rightarrow \mathbb{R}$ is submodular, then so is $\sum_i \alpha_i f_i$. 
Summary so far

- Summing: if $\alpha_i \geq 0$ and $f_i : 2^V \rightarrow \mathbb{R}$ is submodular, then so is $\sum_i \alpha_i f_i$.
- Restrictions: $f'(A) = f(A \cap S)$
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.
Summary so far

- Summing: if $\alpha_i \geq 0$ and $f_i : 2^V \rightarrow \mathbb{R}$ 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 $f(A) = \log \det(\Sigma_A)$
Let $m \in \mathbb{R}_+^E$ be a non-negative modular function, and $g$ a concave function over $\mathbb{R}$. Define $f : 2^E \rightarrow \mathbb{R}$ as

$$f(A) = g(m(A))$$

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

$$g(m(A) + m(v)) - g(m(A)) \geq g(m(B) + m(v)) - g(m(B))$$  \hspace{1cm} (4.43)

A form of converse is true as well.
Theorem 4.5.1

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

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

2. \( g : \mathbb{R}_+ \rightarrow \mathbb{R} \) is concave.

If \( g \) is non-decreasing concave, then \( f \) is polymatroidal.

\[ g(0) = 0 \]
Theorem 4.5.1

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

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

2. $g : \mathbb{R}_+ \rightarrow \mathbb{R}$ is concave.

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

$$f(A) = \sum_{i=1}^{K} g_i(m_i(A))$$ (4.44)
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

$$f(A) = \sum_{i=1}^{K} g_i(m_i(A)) \quad (4.44)$$

- Very large class of functions, including graph cut, bipartite neighborhoods, set cover (Stobbe & Krause 2011), and “feature-based submodular functions” (Wei, Iyer, & Bilmes 2014).
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

\[
    f(A) = \sum_{i=1}^{K} g_i(m_i(A))
\]  

(4.44)

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

**Definition 4.5.2**

A function \( f : 2^V \rightarrow \mathbb{R} \) is **monotone nondecreasing** (resp. **monotone increasing**) if for all \( A \subseteq B \), we have \( f(A) \leq f(B) \) (resp. \( f(A) < f(B) \)).

**Definition 4.5.3**

A function \( f : 2^V \rightarrow \mathbb{R} \) is **monotone nonincreasing** (resp. **monotone decreasing**) if for all \( A \subseteq B \), 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*

\[ f : 2^V \rightarrow \mathbb{R} \quad (4.45) \]

*and another continuous valued one:*

\[ g : \mathbb{R} \rightarrow \mathbb{R} \quad (4.46) \]

*the composition formed as* \( h = g \circ f : 2^V \rightarrow \mathbb{R} \) *(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

\[
h(A) = \min(f(A), g(A))
\]

is submodular.

Proof.

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

\[
f(X) + f(Y) \geq f(X \cup Y) + f(X \cap Y)
\]

\[
g(X) + g(Y) \geq g(X \cup Y) + g(X \cap Y),
\]

the result (Equation 4.47 being submodular) follows since

\[
\frac{f(X) + f(Y)}{g(X) + g(Y)} \geq \min(f(X \cup Y), g(X \cup Y)) + \min(f(X \cap Y), g(X \cap Y))
\]

(4.50)
Otherwise, w.l.o.g., \( h(X) = f(X) \) and \( h(Y) = g(Y) \), giving

\[
h(X) + h(Y) = f(X) + g(Y) \geq f(X \cup Y) + f(X \cap Y) + g(Y) - f(Y)
\]  
(4.51)
Monotone difference of two functions

...cont.

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

\[
h(X) + h(Y) = f(X) + g(Y) \geq f(X \cup Y) + f(X \cap Y) + g(Y) - f(Y)
\]  

(4.51)

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

giving

\[
h(X) + h(Y) \geq g(X \cup Y) + f(X \cap Y) \geq h(X \cup Y) + h(X \cap Y)
\]  

(4.52)

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

$$h(A) = \min(k, f(A))$$

is submodular.
Saturation via the $\min(\cdot)$ function

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

$$h(A) = \min(k, f(A))$$

is submodular.

**Proof.**

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

$$h(A) = \min(k, f(A))$$

(4.53)

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.
In general, the minimum of two submodular functions is not submodular (unlike concave functions, closed under min).
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

$$h_\alpha(A) = \min(\alpha, f(A)) + \min(\alpha, g(A))$$

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 $\alpha \in \mathbb{R}$. 

(4.54)
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

$$h_\alpha(A) = \min(\alpha, f(A)) + \min(\alpha, g(A)) \quad (4.54)$$

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:

$$\alpha \triangleq \min_{X,Y} \left( f(X) + f(Y) - f(X \cup Y) - f(X \cap Y) \right) \quad (4.55)$$

If $\alpha \geq 0$ then $f$ is submodular, so by assumption $\alpha < 0$. 
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:

$$\alpha \triangleq \min_{X,Y} \left( f(X) + f(Y) - f(X \cup Y) - f(X \cap Y) \right)$$  \hspace{1cm} (4.55)

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

$$\beta \triangleq \min_{X,Y: X \not\subseteq Y, Y \not\subseteq X} \left( h(X) + h(Y) - h(X \cup Y) - h(X \cap Y) \right).$$  \hspace{1cm} (4.56)

Strict means that $\beta > 0$. 

...
Arbitrary functions as difference between submodular funcs.

...cont.

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

$$f'(A) = f(A) + \frac{|\alpha|}{\beta} h(A)$$  \hspace{1cm} (4.57)

Then $f'$ is submodular (why?), and $f = f'(A) - \frac{|\alpha|}{\beta} h(A)$, 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 $f(A \cup \{j\}) - f(A)$. 

\[ f(A \cup \{j\}) - f(A) \]
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:

\[
\begin{align*}
    f(A \cup \{j\}) - f(A) & \triangleq \rho_j(A) \quad (4.58) \\
    & \triangleq \rho_A(j) \quad (4.59) \\
    & \triangleq \nabla_j f(A) \quad (4.60) \\
    & \triangleq f(\{j\} | A) \quad (4.61) \\
    & \triangleq f(j | A) \quad (4.62)
\end{align*}
\]
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:

$$f(A \cup \{j\}) - f(A) \triangleq \rho_j(A) \quad (4.58)$$

$$\triangleq \rho_A(j) \quad (4.59)$$

$$\triangleq \nabla_j f(A) \quad (4.60)$$

$$\triangleq f(\{j\} | A) \quad (4.61)$$

$$\triangleq f(j | A) \quad (4.62)$$

- We’ll use $f(j | A)$. 
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:

$$f(A \cup \{j\}) - f(A) \triangleq \rho_j(A)$$  \hspace{1cm} (4.58)

$$\triangleq \rho_A(j)$$  \hspace{1cm} (4.59)

$$\triangleq \nabla_j f(A)$$  \hspace{1cm} (4.60)

$$\triangleq f(\{j\} | A)$$  \hspace{1cm} (4.61)

$$\triangleq f(j | A)$$  \hspace{1cm} (4.62)

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

\[ f(A|B) \triangleq f(A \cup B) - f(B) \]  

(4.63)

So when $j$ is any singleton

\[ f(j|B) = f\{j\}|B = f\{j\} \cup B) - f(B) \]  

(4.64)
Gain Notation

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

\[
 f(A|B) \triangleq f(A \cup B) - f(B) \quad (4.63)
\]

So when \( j \) is any singleton

\[
 f(j|B) = f(\{j\}|B) = f(\{j\} \cup B) - f(B) \quad (4.64)
\]

Note that this is inspired from information theory and the notation used for conditional entropy \( H(X_A|X_B) = H(X_A, X_B) - H(X_B) \).
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 $m_g$. 
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 $g : 2^V \rightarrow \mathbb{R}$, construct $\bar{g} : 2^V \rightarrow \mathbb{R}$ 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,
  $$f = g + h = \bar{g} + m_g + \bar{h} + m_h = \bar{g} + m_g + \bar{h} + m_h$$
  where $m_h$ is the positive part of modular function $m_h$. This, $m_h(A) = \sum_{a \in A} m_h(a)$.

- But both $\bar{g} + m_g + \bar{h} + m_h$ and $\bar{h} + (m_g + m_h)$ are polymatroid functions.

- Thus, any function can be expressed as a difference between two, not only submodular (DS), but polymatroid functions.
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,

$$f = g - h = \bar{g} + m_g - (\bar{h} + m_h) \quad (4.65)$$

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

$$= \bar{g} - \bar{h} + m_{g-h} \quad (4.67)$$

$$= \bar{g} + m^+_{g-h} - (\bar{h} + (-m_{g-h})^+) \quad (4.68)$$

where $m^+$ is the positive part of modular function $m$. That is, $m^+(A) = \sum_{a \in A} m(a)1(m(a) > 0)$. 

Prof. Jeff Bilmes
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

$$
\bar{g}(A) = g(A) - \sum_{a \in A} g(a\mid V \setminus \{a\}).
$$

Let $m_g(A) \triangleq \sum_{a \in A} g(a\mid V \setminus \{a\})$.

Then, given arbitrary $f = g - h$ where $g$ and $h$ are normalized submodular,

$$f = g - h = \bar{g} + m_g - (\bar{h} + m_h)$$

$$= \bar{g} - \bar{h} + (m_g - m_h)$$

$$= \bar{g} - \bar{h} + m_{g-h}$$

$$= \bar{g} + m_{g-h}^+ - (\bar{h} + (-m_{g-h})^+)$$

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.
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,
\[
f = g - h = \bar{g} + m_g - (\bar{h} + m_h) \tag{4.65}
\]
\[
= \bar{g} - \bar{h} + (m_g - m_h) \tag{4.66}
\]
\[
= \bar{g} - \bar{h} + m_{g-h} \tag{4.67}
\]
\[
= \bar{g} + m_{g-h}^+ - (\bar{h} + (-m_{g-h})^+) \tag{4.68}
\]

where $m^+$ is the positive part of modular function $m$. That is,
\[
m^+(A) = \sum_{a \in A} m(a) 1(m(a) > 0).
\]

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
Two Equivalent **Submodular** Definitions

**Definition 4.6.1 (submodular concave)**

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

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

(4.8)

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

**Definition 4.6.2 (diminishing returns)**

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

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

(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 \( f : 2^V \rightarrow \mathbb{R} \) is submodular if for any \( A \subseteq B \subseteq V \), and \( C \subseteq V \setminus B \), we have that:

\[
    f(A \cup C) - f(A) \geq f(B \cup C) - f(B)
\]  

(4.69)

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)
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.
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
  \[
f(S) + f(T) \geq f(S \cup T) + f(S \cap T).
\]
Submodular Concave \implies 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:

  \[
  f(A + v) + f(B) \geq f(B + v) + f(A) \tag{4.70}
  \]
**Submodular Concave ⇒ 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:
  \[
  f(A + v) + f(B) \geq f(B + v) + f(A) \quad (4.70)
  \]
- Rearranging, we have
  \[
  f(A + v) - f(A) \geq f(B + v) - f(B) \quad (4.71)
  \]
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

\[
f(A \cup C) - f(A) = f(A \cup C) - \sum_{i=1}^{k-1} \left( f(A \cup \{c_1, \ldots, c_i\}) - f(A \cup \{c_1, \ldots, c_i\}) \right) - f(A)
\]

\[
\geq \sum_{i=1}^{k} \left( f(B \cup \{c_1 \ldots c_i\}) - f(B \cup \{c_1 \ldots c_{i-1}\}) \right)
\]

\[
= f(B \cup C) - \sum_{i=1}^{k-1} \left( f(B \cup \{c_1, \ldots, c_i\}) - f(B \cup \{c_1, \ldots, c_i\}) \right) - f(B)
\]

\[
= f(B \cup C) - f(B)
\]
Group Diminishing Returns ⇒ Submodular Concave

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

\[
f(A' + C) - f(A') \geq f(B' + C) - f(B')
\]

(4.78)

giving

\[
f(A' + C) + f(B') \geq f(B' + C) + f(A')
\]

(4.79)

or

\[
f(A \cap B + A \setminus B) + f(B) \geq f(B + A \setminus B) + f(A \cap B)
\]

(4.80)

which is the same as the submodular concave condition

\[
f(A) + f(B) \geq f(A \cup B) + f(A \cap B)
\]

(4.81)
Submodular Definition: Four Points

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

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

\[
 f(A \cup \{a\}) + f(A \cup \{b\}) \geq f(A \cup \{a, b\}) + f(A)
\]  

(4.82)
Definition 4.6.2 ("singleton", or "four points")

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

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

This follows immediately from diminishing returns.
Submodular Definition: Four Points

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

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

$$f(A \cup \{a\}) + f(A \cup \{b\}) \geq f(A \cup \{a, b\}) + f(A)$$  \hspace{1cm} (4.82)

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

$$f(A + a) - f(A) \geq f(A + b_1 + a) - f(A + b_1)$$  \hspace{1cm} (4.83)

$$\geq f(A + b_1 + b_2 + a) - f(A + b_1 + b_2)$$  \hspace{1cm} (4.84)

$$\geq \ldots$$  \hspace{1cm} (4.85)

$$\geq f(A + b_1 + \cdots + b_k + a) - f(A + b_1 + \cdots + b_k)$$  \hspace{1cm} (4.86)

$$= f(B + a) - f(B)$$  \hspace{1cm} (4.87)
Submodular on Hypercube Vertices

- Test submodularity via values on vertices of hypercube.

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

$$
\begin{array}{ccc}
00 & 01 & 11 \\
10 & & \\
00 & 01 & \\
\end{array}
$$

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

$$
\begin{array}{cccccc}
000 & 001 & 010 & 011 & 100 & 101 \\
110 & 111 & 101 & 110 & & \\
000 & 001 & 010 & 011 & & \\
\end{array}
$$

How many inequalities?
Submodular Definitions

Theorem 4.6.3

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

\[
f(A) + f(B) \geq f(A \cup B) + f(A \cap B) \quad \text{for all } A, B \subseteq V \quad \text{(SC)}
\]

if and only if

\[
f(v|X) \geq f(v|Y) \quad \text{for all } X \subseteq Y \subseteq V \text{ and } v \notin Y \quad \text{(DR)}
\]

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:

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

\[
\Rightarrow \quad f(A) - f(A \cap B) \geq f(A \cup B) - f(B)
\]
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:

$$f(C) - f(D)$$  \hfill (4.88)
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:

$$f(C) - f(D)$$  \hspace{1cm} (4.88)

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

$$f(C) - f(D) = f(C) - f(C \cup X) \geq f(C \cap X) - f(X)$$  \hspace{1cm} (4.90)
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:

$$f(C) - f(D) \quad (4.88)$$

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

$$f(C) - f(D) = f(C) - f(C \cup X) \geq f(C \cap X) - f(X) \quad (4.89)$$

or

$$f(C \cup X|C) \leq f(X|C \cap X) \quad (4.90)$$
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:
  \[ f(C) - f(D) \]  \hspace{1cm} (4.88)

- If $D \supseteq C$, then for any $X$ with $D = C \cup X$ then
  \[ f(C) - f(D) = f(C) - f(C \cup X) \geq f(C \cap X) - f(X) \]  \hspace{1cm} (4.89)
  or
  \[ f(C \cup X | C) \leq f(X | C \cap X) \]  \hspace{1cm} (4.90)

- Alternatively, if $D \subseteq C$, given any $Y$ such that $D = C \cap Y$ then
  \[ f(C) - f(D) = f(C) - f(C \cap Y) \geq f(C \cup Y) - f(Y) \]  \hspace{1cm} (4.92)
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:

\[
    f(C) - f(D)
\]

(4.88)

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

\[
    f(C) - f(D) = f(C) - f(C \cup X) \geq f(C \cap X) - f(X)
\]

(4.89)

or

\[
    f(C \cup X|C) \leq f(X|C \cap X)
\]

(4.90)

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

\[
    f(C) - f(D) = f(C) - f(C \cap Y) \geq f(C \cup Y) - f(Y)
\]

(4.91)

or

\[
    f(C|C \cap Y) \geq f(C \cup Y|Y)
\]

(4.92)
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:

$$f(C) - f(D)$$ (4.88)

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

$$f(C) - f(D) = f(C) - f(C \cup X) \geq f(C \cap X) - f(X)$$ (4.89)

or

$$f(C \cup X|C) \leq f(X|C \cap X)$$ (4.90)

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

$$f(C) - f(D) = f(C) - f(C \cap Y) \geq f(C \cup Y) - f(Y)$$ (4.91)

or

$$f(C|C \cap Y) \geq f(C \cup Y|Y)$$ (4.92)

- Equations (4.90) and (4.92) have same form.
Many (Equivalent) Definitions of Submodularity

\[ f(A) + f(B) \geq f(A \cup B) + f(A \cap B), \quad \forall A, B \subseteq V \quad (4.93) \]
Many (Equivalent) Definitions of Submodularity

\[ f(A) + f(B) \geq f(A \cup B) + f(A \cap B), \quad \forall A, B \subseteq V \quad (4.93) \]

\[ f(j|S) \geq f(j|T), \quad \forall S \subseteq T \subseteq V, \text{ with } j \in V \setminus T \quad (4.94) \]
Many (Equivalent) Definitions of Submodularity

\[ f(A) + f(B) \geq f(A \cup B) + f(A \cap B), \; \forall A, B \subseteq V \]  \hspace{1cm} (4.93)

\[ f(j|S) \geq f(j|T), \; \forall S \subseteq T \subseteq V, \; \text{with } j \in V \setminus T \]  \hspace{1cm} (4.94)

\[ f(C|S) \geq f(C|T), \; \forall S \subseteq T \subseteq V, \; \text{with } C \subseteq V \setminus T \]  \hspace{1cm} (4.95)
Many (Equivalent) Definitions of Submodularity

\[ f(A) + f(B) \geq f(A \cup B) + f(A \cap B), \ \forall A, B \subseteq V \]  \hspace{1cm} (4.93)

\[ f(j|S) \geq f(j|T), \ \forall S \subseteq T \subseteq V, \ \text{with } j \in V \setminus T \]  \hspace{1cm} (4.94)

\[ f(C|S) \geq f(C|T), \ \forall S \subseteq T \subseteq V, \ \text{with } C \subseteq V \setminus T \]  \hspace{1cm} (4.95)

\[ f(j|S) \geq f(j|S \cup \{k\}), \ \forall S \subseteq V \ \text{with } j \in V \setminus (S \cup \{k\}) \]  \hspace{1cm} (4.96)
Many (Equivalent) Definitions of Submodularity

\[ f(A) + f(B) \geq f(A \cup B) + f(A \cap B), \quad \forall A, B \subseteq V \quad (4.93) \]

\[ f(j|S) \geq f(j|T), \quad \forall S \subseteq T \subseteq V, \text{ with } j \in V \setminus T \quad (4.94) \]

\[ f(C|S) \geq f(C|T), \quad \forall S \subseteq T \subseteq V, \text{ with } C \subseteq V \setminus T \quad (4.95) \]

\[ f(j|S) \geq f(j|S \cup \{k\}), \quad \forall S \subseteq V \text{ with } j \in V \setminus (S \cup \{k\}) \quad (4.96) \]

\[ f(A \cup B|A \cap B) \leq f(A|A \cap B) + f(B|A \cap B), \quad \forall A, B \subseteq V \quad (4.97) \]
Many (Equivalent) Definitions of Submodularity

\[ f(A) + f(B) \geq f(A \cup B) + f(A \cap B), \quad \forall A, B \subseteq V \]  
(4.93)

\[ f(j|S) \geq f(j|T), \quad \forall S \subseteq T \subseteq V, \text{ with } j \in V \setminus T \]  
(4.94)

\[ f(C|S) \geq f(C|T), \quad \forall S \subseteq T \subseteq V, \text{ with } C \subseteq V \setminus T \]  
(4.95)

\[ f(j|S) \geq f(j|S \cup \{k\}), \quad \forall S \subseteq V \text{ with } j \in V \setminus (S \cup \{k\}) \]  
(4.96)

\[ f(A \cup B|A \cap B) \leq f(A|A \cap B) + f(B|A \cap B), \quad \forall A, B \subseteq V \]  
(4.97)

\[ f(T) \leq f(S) + \sum_{j \in T \setminus S} f(j|S) - \sum_{j \in S \setminus T} f(j|S \cup T - \{j\}), \quad \forall S, T \subseteq V \]  
(4.98)
Many (Equivalent) Definitions of Submodularity

\[
f(A) + f(B) \geq f(A \cup B) + f(A \cap B), \quad \forall A, B \subseteq V \tag{4.93}
\]
\[
f(j | S) \geq f(j | T), \quad \forall S \subseteq T \subseteq V, \text{ with } j \in V \setminus T \tag{4.94}
\]
\[
f(C | S) \geq f(C | T), \quad \forall S \subseteq T \subseteq V, \text{ with } C \subseteq V \setminus T \tag{4.95}
\]
\[
f(j | S) \geq f(j | S \cup \{k\}), \quad \forall S \subseteq V \text{ with } j \in V \setminus (S \cup \{k\}) \tag{4.96}
\]
\[
f(A \cup B | A \cap B) \leq f(A | A \cap B) + f(B | A \cap B), \quad \forall A, B \subseteq V \tag{4.97}
\]
\[
f(T) \leq f(S) + \sum_{j \in T \setminus S} f(j | S) - \sum_{j \in S \setminus T} f(j | S \cup T - \{j\}), \quad \forall S, T \subseteq V \tag{4.98}
\]
\[
f(T) \leq f(S) + \sum_{j \in T \setminus S} f(j | S), \quad \forall S \subseteq T \subseteq V \tag{4.99}
\]
Many (Equivalent) Definitions of Submodularity

\[ f(A) + f(B) \geq f(A \cup B) + f(A \cap B), \quad \forall A, B \subseteq V \quad (4.93) \]
\[ f(j|S) \geq f(j|T), \quad \forall S \subseteq T \subseteq V, \text{ with } j \in V \setminus T \quad (4.94) \]
\[ f(C|S) \geq f(C|T), \quad \forall S \subseteq T \subseteq V, \text{ with } C \subseteq V \setminus T \quad (4.95) \]
\[ f(j|S) \geq f(j|S \cup \{k\}), \quad \forall S \subseteq V \text{ with } j \in V \setminus (S \cup \{k\}) \quad (4.96) \]
\[ f(A \cup B|A \cap B) \leq f(A|A \cap B) + f(B|A \cap B), \quad \forall A, B \subseteq V \quad (4.97) \]
\[ f(T) \leq f(S) + \sum_{j \in T \setminus S} f(j|S) - \sum_{j \in S \setminus T} f(j|S \cup T - \{j\}), \quad \forall S, T \subseteq V \quad (4.98) \]
\[ f(T) \leq f(S) + \sum_{j \in T \setminus S} f(j|S), \quad \forall S \subseteq T \subseteq V \quad (4.99) \]
\[ f(T) \leq f(S) - \sum_{j \in S \setminus T} f(j|S \setminus \{j\}) + \sum_{j \in T \setminus S} f(j|S \cap T) \quad \forall S, T \subseteq V \quad (4.100) \]
Many (Equivalent) Definitions of Submodularity

\[
f(A) + f(B) \geq f(A \cup B) + f(A \cap B), \; \forall A, B \subseteq V \tag{4.93}
\]
\[
f(j|S) \geq f(j|T), \; \forall S \subseteq T \subseteq V, \text{ with } j \in V \setminus T \tag{4.94}
\]
\[
f(C|S) \geq f(C|T), \; \forall S \subseteq T \subseteq V, \text{ with } C \subseteq V \setminus T \tag{4.95}
\]
\[
f(j|S) \geq f(j|S \cup \{k\}), \; \forall S \subseteq V \text{ with } j \in V \setminus (S \cup \{k\}) \tag{4.96}
\]
\[
f(A \cup B|A \cap B) \leq f(A|A \cap B) + f(B|A \cap B), \; \forall A, B \subseteq V \tag{4.97}
\]
\[
f(T) \leq f(S) + \sum_{j \in T \setminus S} f(j|S) - \sum_{j \in S \setminus T} f(j|S \cup T - \{j\}), \; \forall S, T \subseteq V \tag{4.98}
\]
\[
f(T) \leq f(S) + \sum_{j \in T \setminus S} f(j|S), \; \forall S \subseteq T \subseteq V \tag{4.99}
\]
\[
f(T) \leq f(S) - \sum_{j \in S \setminus T} f(j|S \setminus \{j\}) + \sum_{j \in T \setminus S} f(j|S \cap T) \; \forall S, T \subseteq V \tag{4.100}
\]
\[
f(T) \leq f(S) - \sum_{j \in S \setminus T} f(j|S \setminus \{j\}), \; \forall T \subseteq S \subseteq V \tag{4.101}
\]
We’ve already seen that Eq. 4.93 \equiv Eq. 4.94 \equiv Eq. 4.95 \equiv Eq. 4.96 \equiv Eq. 4.97.
Equivalent Definitions of Submodularity

We’ve already seen that Eq. 4.93 $\equiv$ Eq. 4.94 $\equiv$ Eq. 4.95 $\equiv$ Eq. 4.96 $\equiv$ Eq. 4.97.

We next show that Eq. 4.96 $\Rightarrow$ Eq. 4.98 $\Rightarrow$ Eq. 4.99 $\Rightarrow$ Eq. 4.96.
To show these next results, we essentially first use:

\[
f(S \cup T) = f(S) + f(T|S) \leq f(S) + \text{upper-bound} \quad (4.102)
\]

and

\[
f(T) + \text{lower-bound} \leq f(T) + f(S|T) = f(S \cup T) \quad (4.103)
\]
Approach

To show these next results, we essentially first use:

\[ f(S \cup T) = f(S) + f(T|S) \leq f(S) + \text{upper-bound} \quad (4.102) \]

and

\[ f(T) + \text{lower-bound} \leq f(T) + f(S|T) = f(S \cup T) \quad (4.103) \]

leading to

\[ f(T) + \text{lower-bound} \leq f(S) + \text{upper-bound} \quad (4.104) \]

or

\[ f(T) \leq f(S) + \text{upper-bound} - \text{lower-bound} \quad (4.105) \]
Eq. 4.96 $\Rightarrow$ Eq. 4.98

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 $S$:

\[
\begin{align*}
\quad f(\overline{S} \cup T) - f(\overline{S}) &= \sum_{t=1}^{r} \left( f(\overline{S} \cup \{j_1, \ldots, j_t\}) - f(\overline{S} \cup \{j_1, \ldots, j_{t-1}\}) \right) \\
&= \sum_{t=1}^{r} f(j_t | S \cup \{j_1, \ldots, j_{t-1}\}) \leq \sum_{t=1}^{r} f(j_t | S) \\
&= \sum_{j \in T \setminus S} f(j | S) \\
\end{align*}
\]

(4.106)

(4.107)

(4.108)

or

\[
\begin{align*}
\quad f(T | S) &\leq \sum_{j \in T \setminus S} f(j | S) \\
\end{align*}
\]

(4.109)
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 $T$:

\[
\begin{align*}
\quad & f(S \cup T) - f(T) = \sum_{t=1}^{q} \left[ f(T \cup \{k_1, \ldots, k_t\}) - f(T \cup \{k_1, \ldots, k_{t-1}\}) \right] \\
\quad & = \sum_{t=1}^{q} f(k_t | T \cup \{k_1, \ldots, k_t\} \setminus \{k_t\}) \geq \sum_{t=1}^{q} f(k_t | T \cup S \setminus \{k_t\}) \\
\quad & = \sum_{j \in S \setminus T} f(j | S \cup T \setminus \{j\})
\end{align*}
\]
Let $T \setminus S = \{j_1, \ldots, j_r\}$ and $S \setminus T = \{k_1, \ldots, k_q\}$.

So we have the upper bound

$$f(T|S) = f(S \cup T) - f(S) \leq \sum_{j \in T \setminus S} f(j|S)$$  \hspace{1cm} (4.113)

and the lower bound

$$f(S|T) = f(S \cup T) - f(T) \geq \sum_{j \in S \setminus T} f(j|S \cup T \setminus \{j\})$$  \hspace{1cm} (4.114)

This gives upper and lower bounds of the form

$$f(T) + \text{lower bound} \leq f(S \cup T) \leq f(S) + \text{upper bound},$$  \hspace{1cm} (4.115)

and combining directly the left and right hand side gives the desired inequality.
Eq. 4.98 $\Rightarrow$ Eq. 4.99

This follows immediately since if $S \subseteq T$, then $S \setminus T = \emptyset$, and the last term of Eq. 4.98 vanishes.
Here, we set $T = S \cup \{j, k\}$, $j \notin S \cup \{k\}$ into Eq. 4.99 to obtain

$$f(S \cup \{j, k\}) \leq f(S) + f(j|S) + f(k|S) \quad (4.116)$$

$$= f(S) + f(S + \{j\}) - f(S) + f(S + \{k\}) - f(S) \quad (4.117)$$

$$= f(S + \{j\}) + f(S + \{k\}) - f(S) \quad (4.118)$$

$$= f(j|S) + f(S + \{k\}) \quad (4.119)$$

giving

$$f(j|S \cup \{k\}) = f(S \cup \{j, k\}) - f(S \cup \{k\}) \quad (4.120)$$

$$\leq f(j|S) \quad (4.121)$$
Why do we call the $f(A) + f(B) \geq f(A \cup B) + f(A \cap B)$ definition of submodularity, 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 $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is concave iff $\nabla^2 f \preceq 0$ (the Hessian matrix is nonpositive definite).
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 \( \nabla^2 f \leq 0 \) (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:

\[
(\nabla_B f)(A) \overset{\Delta}{=} f(A \cup B) - f(A \setminus B) = f(B|A \setminus B))
\]

(4.122)

read as: the derivative of \( f \) at \( A \) in the direction \( B \).
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 \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) 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 \( f : 2^V \rightarrow \mathbb{R} \) as follows:
  \[
  (\nabla_B f)(A) \overset{\Delta}{=} f(A \cup B) - f(A \setminus B) = f(B | (A \setminus B))
  \]
  \[\text{(4.122)}\]
  read as: the derivative of \( f \) at \( A \) in the direction \( B \).
- Hence, if \( A \cap B = \emptyset \), then \( (\nabla_B f)(A) = f(B | A) \).
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 $\nabla^2 f \leq 0$ (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:

$$\left(\nabla_B f\right)(A) \triangleq f(A \cup B) - f(A \setminus B) = f(B | (A \setminus B))$$  \hspace{1cm} (4.122)

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:

$$\left(\nabla_C \nabla_B f\right)(A) = \nabla_C \left[ f(A \cup B) - f(A \setminus B) \right]$$

$$= (\nabla_B f)(A \cup C) - (\nabla_B f)(A \setminus C)$$

$$= f(A \cup B \cup C) - f((A \cup C) \setminus B)$$

$$- f((A \setminus C) \cup B) + f((A \setminus C) \setminus B)$$  \hspace{1cm} (4.123) (4.124) (4.125)
Submodular Concave

- If the second difference operator everywhere nonpositive:

\[
f(A \cup B \cup C) - f((A \cup C) \setminus B) \\
- f((A \setminus C) \cup B) + f(A \setminus C \setminus B) \leq 0 \tag{4.126}
\]
If the second difference operator everywhere nonpositive:

\[
\begin{align*}
  f(A \cup B \cup C) &- f((A \cup C) \setminus B) \\
  &- f((A \setminus C) \cup B) + f(A \setminus C \setminus B) \leq 0
\end{align*}
\]  \hspace{1cm} (4.126)

then we have the equation:

\[
\begin{align*}
  f((A \cup C) \setminus B) + f((A \setminus C) \cup B) &\geq f(A \cup B \cup C) + f(A \setminus C \setminus B)
\end{align*}
\]  \hspace{1cm} (4.127)
Submodular Concave

• If the second difference operator everywhere nonpositive:

$$f(A \cup B \cup C) - f((A \cup C) \setminus B)$$

$$- f((A \setminus C) \cup B) + f(A \setminus C \setminus B) \leq 0$$  \hspace{1cm} (4.126)

then we have the equation:

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

\hspace{1cm} (4.127)

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

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

\hspace{1cm} (4.128)

and note that \(A'\) and \(B'\) so defined can be arbitrary.
Submodular Concave

- If the second difference operator everywhere nonpositive:

\[
\begin{align*}
    f(A \cup B \cup C') - f((A \cup C') \setminus B) \\
    - f((A \setminus C') \cup B) + f(A \setminus C' \setminus B) & \leq 0
\end{align*}
\]

then we have the equation:

\[
\begin{align*}
    f((A \cup C') \setminus B) + f((A \setminus C') \cup B) & \geq f(A \cup B \cup C') + f(A \setminus C \setminus B)
\end{align*}
\]

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

\[
\begin{align*}
    f(A') + f(B') & \geq f(A' \cup B') + f(A' \cap B')
\end{align*}
\]

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 $A' \cap B' = A \setminus C \setminus B$. 
Figure: A figure showing $A' \cup B' = A \cup B \cup C$ and $A' \cap B' = A \setminus C \setminus B$.
This submodular/concave relationship is more simply done with singletons.
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 $j, k \in V \setminus X$

\[
f(X + j) + f(X + k) \geq f(X + j + k) + f(X)
\]  

(4.129)
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 $j, k \in V \setminus X$

  $$f(X + j) + f(X + k) \geq f(X + j + k) + f(X) \quad (4.129)$$

- This gives us a simpler notion corresponding to concavity.
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 \( j, k \in V \setminus X \)

\[
f(X + j) + f(X + k) \geq f(X + j + k) + f(X)
\]

(4.129)

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.
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 $j, k \in V \setminus X$

$$f(X + j) + f(X + k) \geq f(X + j + k) + f(X)$$  \hspace{1cm} (4.129)

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:

$$\nabla_j \nabla_k f(X) \leq 0$$  \hspace{1cm} (4.130)