Announcements, Assignments, and Reminders

- Reminder: class web page is at our web page (http://j.ee.washington.edu/~bilmes/classes/ee596a_fall_2012/)
- Please do use our discussion board (https://catalyst.uw.edu/gopost/board/bilmes/29948/) for all questions, comments, so that all will benefit from them being answered.
Outstanding Reading

- Read chapter 1 from Fujishige book.
Many Applications

- In the last lecture, we saw instances of many different problems that we might want to solve.
Many Applications

- In the last lecture, we saw instances of many different problems that we might want to solve.
- It turns out that they are all instances of, or closely related to, the concept of submodularity or supermodularity.
Submodular Motivation Recap

- Given a set of objects $V = \{v_1, \ldots, v_n\}$ and a function $f : 2^V \to \mathbb{R}$ that returns a real value for any subset $S \subseteq V$.
- Suppose we are interested in finding the subset that either maximizes or minimizes the function, e.g., $\text{argmax}_{S \subseteq V} f(S)$, possibly subject to some constraints.
- In general, this problem has exponential time complexity.
- Example: $f$ might correspond to the value (e.g., information gain) of a set of sensor locations in an environment, and we wish to find the best set $S \subseteq V$ of sensors locations given a fixed upper limit on the number of sensors $|S|$.
- In many cases (such as above) $f$ has properties that make its optimization tractable to either exactly or approximately compute.
- One such property is submodularity.
**Submodular Definitions**

**Definition 2.2.4 (submodular concave)**

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

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

(2.7)

An alternate and (as we see in lecture 3) equivalent definition is:

**Definition 2.2.5 (diminishing returns)**

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

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

(2.8)

This means that the incremental “value”, “gain”, or “cost” of \( v \) decreases (diminishes) as the context in which \( v \) is considered grows from \( A \) to \( B \).
Supermodular Definitions

Definition 2.2.4 (supermodular convex)

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

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

(2.7)

An alternate and equivalent definition is:

Definition 2.2.5 (increasing returns)

A function \( f: 2^V \rightarrow \mathbb{R} \) is supermodular if for any \( A \subseteq B \subset V \), and \( v \in V \setminus B \), we have that:

\[
f(A \cup \{v\}) - f(A) \leq f(B \cup \{v\}) - f(B)
\]  

(2.8)

The incremental “value”, “gain”, or “cost” of \( v \) increases as the context in which \( v \) is considered grows from \( A \) to \( B \).
Submodular vs. Supermodular

- Submodular and supermodular functions are closely related.
Submodular vs. Supermodular

- Submodular and supermodular functions are closely related.
- In fact, $f$ is submodular iff $-f$ is supermodular.
Modular Definitions

Definition 2.2.5 (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

$$f(A) + f(B) = f(A \cap B) + f(A \cup B)$$  \hspace{1cm} (2.8)

Modular functions have no interaction, and have value based only on singleton values.

Proposition 2.2.6
If $f$ is modular, it may be written as

$$f(A) = f(\emptyset) + \sum_{a \in A} \left( f(\{a\}) - f(\emptyset) \right)$$  \hspace{1cm} (2.9)
Complement function

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

**Proposition 2.2.5**

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

**Proof.**

$$\bar{f}(A) + \bar{f}(B) \geq \bar{f}(A \cup B) + \bar{f}(A \cap B) \quad (2.13)$$

follows from

$$f(V \setminus A) + f(V \setminus B) \geq f(V \setminus (A \cup B)) + f(V \setminus (A \cap B)) \quad (2.14)$$

which is true because $V \setminus (A \cup B) = (V \setminus A) \cap (V \setminus B)$ and $V \setminus (A \cap B) = (V \setminus A) \cup (V \setminus B)$. 

\[\square\]
Submodularity

- Submodular functions have a long history in economics, game theory, combinatorial optimization, electrical networks, and operations research.
- They are gaining importance in machine learning as well (one of our main motivations for offering this course).
- Arbitrary set functions are hopelessly difficult to optimize, while the minimum of submodular functions can be found in polynomial time, and the maximum can be constant-factor approximated in low-order polynomial time.
- Submodular functions share properties in common with both convex and concave functions.
Attractions of Convex Functions

Why do we like Convex Functions? (Quoting Lovász 1983):

1. Convex functions occur in many mathematical models in economy, engineering, and other sciences. Convexity is a very natural property of various functions and domains occurring in such models; quite often the only non-trivial property which can be stated in general.
Attractions of Convex Functions

Why do we like Convex Functions? (Quoting Lovász 1983):

1. Convex functions occur in many mathematical models in economy, engineering, and other sciences. Convexity is a very natural property of various functions and domains occurring in such models; quite often the only non-trivial property which can be stated in general.

2. Convexity is preserved under many natural operations and transformations, and thereby the effective range of results can be extended, elegant proof techniques can be developed as well as unforeseen applications of certain results can be given.
Attractons of Convex Functions

Why do we like Convex Functions? (Quoting Lovász 1983):

1. **Convex functions occur in many mathematical models in economy, engineering, and other sciences. Convexity is a very natural property of various functions and domains occurring in such models; quite often the only non-trivial property which can be stated in general.**

2. **Convexity is preserved under many natural operations and transformations, and thereby the effective range of results can be extended, elegant proof techniques can be developed as well as unforeseen applications of certain results can be given.**

3. **Convex functions and domains exhibit sufficient structure so that a mathematically beautiful and practically useful theory can be developed.**
Attractions of Convex Functions

Why do we like Convex Functions? (Quoting Lovász 1983):

1. **Convex functions occur in many mathematical models in economy, engineering, and other sciences. Convexity is a very natural property of various functions and domains occurring in such models; quite often the only non-trivial property which can be stated in general.**

2. **Convexity is preserved under many natural operations and transformations, and thereby the effective range of results can be extended, elegant proof techniques can be developed as well as unforeseen applications of certain results can be given.**

3. **Convex functions and domains exhibit sufficient structure so that a mathematically beautiful and practically useful theory can be developed.**

4. **There are theoretically and practically (reasonably) efficient methods to find the minimum of a convex function.**
In this course, we wish to demonstrate that submodular functions also possess attractions of these four sorts as well.
Example Submodular: Number of Colors of Balls in Urns

- Consider an urn containing colored balls. Given a set $S$ of balls, $f(S)$ counts the number of distinct colors.
Consider an urn containing colored balls. Given a set $S$ of balls, $f(S)$ counts the number of distinct colors.

Initial value: 2 (colors in urn).
New value with added blue ball: 3

Initial value: 3 (colors in urn).
New value with added blue ball: 3

Submodularity: Incremental value of object diminishes in a larger context (diminishing returns). Thus, $f$ is submodular.
Consider an urn containing colored balls. Given a set $S$ of balls, $f(S)$ counts the number of distinct colors.

- Initial value: 2 (colors in urn).
  New value with added blue ball: 3

- Initial value: 3 (colors in urn).
  New value with added blue ball: 3

**Submodularity:** Incremental Value of Object Diminishes in a Larger Context (diminishing returns).
Consider an urn containing colored balls. Given a set $S$ of balls, $f(S)$ counts the number of distinct colors.

- Initial value: 2 (colors in urn).
- New value with added blue ball: 3

- Initial value: 3 (colors in urn).
- New value with added blue ball: 3

Submodularity: Incremental Value of Object Diminishes in a Larger Context (diminishing returns).

Thus, $f$ is submodular.
Consumer costs are very often submodular.
Consumer costs are very often submodular. For example:

\[ f(\text{hamburger}) + f(\text{fries}) \geq f(\text{hamburger} + \text{fries}) + f(\text{hamburger}) \]
Ex. Submodular: Consumer Costs of Living

- Consumer costs are very often submodular. For example:

\[ f(\text{fries}) + f(\text{burger}) \geq f(\text{fries}) + f(\text{burger}) \]

- Rearranging terms, we can see this as diminishing returns:
Ex. Submodular: Consumer Costs of Living

- Consumer costs are very often submodular. For example:

\[ f(\text{coke}) + f(\text{fries}) \geq f(\text{coke + fries}) \]

- Rearranging terms, we can see this as diminishing returns:

\[ f(\text{coke + fries}) - f(\text{coke}) \geq f(\text{coke}) - f(\text{fries}) \]
Consumer costs are very often submodular. For example:

\[ f(\text{coke, fries, burger}) + f(\text{fries, burger}) \geq f(\text{fries, burger}) + f(\text{coke}) \]

Rearranging terms, we can see this as diminishing returns:

\[ f(\text{coke, fries, burger}) - f(\text{fries, burger}) \geq f(\text{fries, burger}) - f(\text{coke}) \]

This is very common: The additional cost of a coke is, say, free if you add it to fries and a hamburger, but when added just to an order of fries, the coke is not free.
Let $V$ be a set of indices, and each $v \in V$ indexes a given sub-area of some region. Let $\text{area}(v)$ be the area corresponding to item $v$.

Let $f(S) = \bigcup_{s \in S} \text{area}(s)$ be the union of the areas indexed by elements in $A$.

Then $f(S)$ is submodular.
Area of the union of areas indexed by $A$

Union of areas of elements of $A$ is given by:

$$f(A) = f(\{a_1, a_2, a_3, a_4\})$$
Area of the union of areas indexed by $A$

Area of $A$ along with $v$:

$$f(A \cup \{v\}) = f(\{a_1, a_2, a_3, a_4\} \cup \{v\})$$
Area of the union of areas indexed by $A$

Gain (value) of $v$ in context of $A$:

$$f(A \cup \{v\}) - f(A) = f(\{v\})$$

We get full value $f(\{v\})$ in this case since the area of $v$ has no overlap with that of $A$. 
Area of the union of areas indexed by $A$

Area of $A$ once again.

$$f(A) = f\left\{a_1, a_2, a_3, a_4\right\}$$
Area of the union of areas indexed by $A$

Union of areas of elements of $B \supset A$, where $v$ is not included:

$$f(B) \text{ where } v \notin B \text{ and where } A \subseteq B$$
Area of the union of areas indexed by $A$

Area of $B$ now also including $v$:

$$f(B \cup \{v\})$$
Area of the union of areas indexed by $A$

Incremental value of $v$ in the context of $B \supseteq A$.

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

So benefit of $v$ in the context of $A$ is greater than the benefit of $v$ in the context of $B \supseteq A$. 
Entropy is submodular. Let $V$ be the index set of a set of random variables, then the function

$$f(A) = H(X_A) = -\sum_{x_A} p(x_A) \log p(x_A) \quad (2.13)$$

is submodular.

Proof: conditioning reduces entropy. With $A \subseteq B$ and $v \notin B$,

$$H(X_v | X_B) = H(X_{B+v}) - H(X_B) \quad (2.14)$$

$$\leq H(X_{A+v}) - H(X_A) = H(X_v | X_A) \quad (2.15)$$
Alternate Proof: Conditional mutual Information is always non-negative.

Consider the following conditional mutual information quantity:

\[
I(X_{A \setminus B}; X_{B \setminus A} | X_{A \cap B}) = \sum_{x_{A \cup B}} p(x_{A \cup B}) \log \frac{p(x_{A \cup B})p(x_{A \cap B})}{p(x_A)p(x_B)} \geq 0
\]  
(2.16)

then

\[
I(X_{A \setminus B}; X_{B \setminus A} | X_{A \cap B}) = H(X_A) + H(X_B) - H(X_{A \cup B}) - H(X_{A \cap B}) \geq 0
\]  
(2.17)

so entropy satisfies

\[
H(X_A) + H(X_B) \geq H(X_{A \cup B}) + H(X_{A \cap B})
\]  
(2.18)
Example Submodular: Mutual Information

Also, symmetric mutual information is submodular,

\[
f(A) = I(X_A; X_{V\setminus A}) = H(X_A) + H(X_{V\setminus A}) - H(X_V)
\]  

(2.19)

Note that \(f(A) = H(X_A)\) and \(\bar{f}(A) = H(X_{V\setminus A})\), and adding submodular functions preserves submodularity (which we will see quite soon).
Undirected Graphs

- Let $G = (V, E)$ be a graph with vertices $V = V(G)$ and edges $E = E(G) \subseteq V \times V$. 

```plaintext
S = \{a, b, c\}
\delta_G(S) = \{\{u \cap v\} \in E : u \in S \cap v \in V \cap S\}
```
Undirected Graphs

- Let $G = (V, E)$ be a graph with vertices $V = V(G)$ and edges $E = E(G) \subseteq V \times V$.
- If $G$ is undirected, define

$$E(X, Y) = \{\{x, y\} \in E(G) : x \in X \setminus Y, y \in Y \setminus X\} \quad (2.20)$$

as the edges between $X$ and $Y$.
Undirected Graphs

- Let $G = (V, E)$ be a graph with vertices $V = V(G)$ and edges $E = E(G) \subseteq V \times V$.
- If $G$ is undirected, define
  \[ E(X, Y) = \{\{x, y\} \in E(G) : x \in X \setminus Y, y \in Y \setminus X\} \]  
  as the edges between $X$ and $Y$.
- Nodes define cuts, and define $\delta(X) = E(X, V \setminus X)$. 

\[ (2.20) \]
Undirected Graphs

- Let $G = (V, E)$ be a graph with vertices $V = V(G)$ and edges $E = E(G) \subseteq V \times V$.
- If $G$ is undirected, define
  
  $$E(X, Y) = \{\{x, y\} \in E(G) : x \in X \setminus Y, y \in Y \setminus X\} \quad (2.20)$$

  as the edges between $X$ and $Y$.
- Nodes define cuts, and define $\delta(X) = E(X, V \setminus X)$. For example:

  $G = (V, E)$

  $\delta_G(S) = \{\{u, v\} \in E : u \in S, v \in V \setminus S\} = \{\{a, d\}, \{b, d\}, \{b, e\}, \{c, e\}, \{c, f\}\}$
Directed Graphs

- If $G$ is directed, define

$$E^+(X, Y) \triangleq \{(x, y) \in E(G) : x \in X \setminus Y, y \in Y \setminus X\} \quad (2.21)$$

as the edges directed from $X$ towards $Y$. 

Nodes define cuts. Define edges leaving $X$ as

$$\delta^+(X) \triangleq E^+(X, V \setminus X) \quad (2.22)$$

and edges entering $X$ as

$$\delta^-(X) \triangleq E^+(V \setminus X, X) \quad (2.23)$$
Directed Graphs

- If $G$ is directed, define
  \[ E^+(X, Y) \triangleq \{ (x, y) \in E(G) : x \in X \setminus Y, y \in Y \setminus X \} \]  
  (2.21)
  as the edges directed from $X$ towards $Y$.

- Nodes define cuts. Define edges leaving $X$ as
  \[ \delta^+(X) \triangleq E^+(X, V \setminus X) \]  
  (2.22)
  and edges entering $X$ as
  \[ \delta^-(X) \triangleq E^+(V \setminus X, X) \]  
  (2.23)
Directed Graphs

If $G$ is directed, define

$$E^+(X, Y) \triangleq \{(x, y) \in E(G) : x \in X \setminus Y, y \in Y \setminus X\}$$  \hspace{1cm} (2.21)

as the edges directed from $X$ towards $Y$.

Nodes define cuts. Define edges leaving $X$ as

$$\delta^+(X) \triangleq E^+(X, V \setminus X)$$  \hspace{1cm} (2.22)

and edges entering $X$ as

$$\delta^-(X) \triangleq E^+(V \setminus X, X)$$  \hspace{1cm} (2.23)

$$\delta_G(S) = \{(v, u) \in E : u \in S, v \in V \setminus S\}$$
= \{(d, a), (d, b), (e, c)\}

$$\delta^+_G(S) = \{(u, v) \in E : u \in S, v \in V \setminus S\}.$$
= \{(b, e), (c, f)\}
Neighborhood function in undirected graphs

- Given a set \( X \subseteq V \), the neighborhood function of \( X \) is defined as

\[
\Gamma(X) \triangleq \{ v \in V(G) \setminus X : E(X, \{v\}) \neq \emptyset \}
\]  

(2.24)
Directed Cut functions

Lemma 2.4.1

For a digraph $G = (V, E)$ and any $X, Y \subseteq V$: we have

$$|\delta^+(X)| + |\delta^+(Y)| = |\delta^+(X \cap Y)| + |\delta^+(X \cup Y)| + |E^+(X, Y)| + |E^+(Y, X)|$$  \hspace{1cm} (2.25)

and

$$|\delta^-(X)| + |\delta^-(Y)| = |\delta^-(X \cap Y)| + |\delta^-(X \cup Y)| + |E^-(X, Y)| + |E^-(Y, X)|$$  \hspace{1cm} (2.26)
Directed Cut functions

Proof.

We can prove this using a simple geometric counting argument ($\delta^-(X)$ is similar)

\[
\begin{align*}
|\delta^+(X)| & \quad |\delta^+(Y)| \\
|\delta^+(X \cap Y)| & \quad |\delta^+(X \cup Y)| \\
|E^+(X, Y)| & \quad |E^+(Y, X)|
\end{align*}
\]
Directed Cut functions

Lemma 2.4.2

For a digraph $G = (V, E)$ and any $X, Y \subseteq V$: both functions $|\delta^+(X)|$ and $|\delta^-(X)|$ are submodular.

Proof.

$|E^+(X, Y)| \geq 0$ and $|E^-(X, Y)| \geq 0$.

Non-negative weighted case: flow is submodular in the vertices.
Lemma 2.4.3

For an undirected graph $G = (V, E)$ and any $X, Y \subseteq V$: we have

\[
|\delta(X)| + |\delta(Y)| = |\delta(X \cap Y)| + |\delta(X \cup Y)| + 2|E(X, Y)| \tag{2.27}
\]

\[
|\Gamma(X)| + |\Gamma(Y)| \geq |\Gamma(X \cap Y)| + |\Gamma(X \cup Y)| \tag{2.28}
\]

Proof.

Eq. (2.27) directly follows from Eq. (2.25) by replacing each edge $\{u, v\}$ with two oppositely directed edges $(u, v)$ and $(v, u)$ and using the same counting argument.

Eq. (2.28) follows since

\[
|\Gamma(X)| + |\Gamma(Y)| = |\Gamma(X \cup Y)| + |\Gamma(X) \cap \Gamma(Y)| + |\Gamma(X) \cap Y| + |\Gamma(Y) \cap X| \geq |\Gamma(X \cap Y)| + |\Gamma(X \cup Y)|
\]

...
Graphically, we can count and see that

\[
\Gamma(X) = (a) + (c) + (f) + (g) + (d) \\
\Gamma(Y) = (b) + (c) + (e) + (h) + (d) \\
\Gamma(X \cup Y) = (a) + (b) + (c) + (d) \\
\Gamma(X \cap Y) = (c) + (g) + (h)
\]

So

\[
|\Gamma(X)| + |\Gamma(Y)| = (a) + (b) + 2(c) + 2(d) + (e) + (f) + (g) + (h) \\
\geq (a) + (b) + 2(c) + (d) + (g) + (h) = |\Gamma(X \cup Y)| + |\Gamma(X \cap Y)|
\]
Therefore, the undirected cut function $|\delta(A)|$ and the neighbor function $|\Gamma(A)|$ of a graph $G$ are both submodular.
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.
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.
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.
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. Thus, we see that $I(S) = E(S) \cup \delta(S)$ and $|I(S)| = |E(S)| + |\delta(S)|$. So we can get a submodular function by summing a submodular and a supermodular function.
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. Thus, we see that $I(S) = E(S) \cup \delta(S)$ and $|I(S)| = |E(S)| + |\delta(S)|$. 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?
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. Thus, we see that $I(S) = E(S) \cup \delta(S)$ and $|I(S)| = |E(S)| + |\delta(S)|$. 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.
Other graph functions that are submodular/supermodular

- Recall, \( f : 2^V \to \mathbb{R} \) is submodular, then so is \( \bar{f} : 2^V \to \mathbb{R} \) defined as \( \bar{f}(S) = f(V \setminus S) \).
Other graph functions that are submodular/supermodular

- Recall, $f : 2^V \rightarrow \mathbb{R}$ is submodular, then so is $\bar{f} : 2^V \rightarrow \mathbb{R}$ defined as $\bar{f}(S) = f(V \setminus S)$.
- Hence, if $f : 2^V \rightarrow \mathbb{R}$ is supermodular, then so is $\bar{f} : 2^V \rightarrow \mathbb{R}$ defined as $\bar{f}(S) = f(V \setminus S)$. 
Recall, $f : 2^V \to \mathbb{R}$ is submodular, then so is $\overline{f} : 2^V \to \mathbb{R}$ defined as $\overline{f}(S) = f(V \setminus S)$.

Hence, if $f : 2^V \to \mathbb{R}$ is supermodular, then so is $\overline{f} : 2^V \to \mathbb{R}$ defined as $\overline{f}(S) = f(V \setminus S)$.

Given a graph, for each $X \subseteq E(G)$, let $c(X)$ denote the number of connected components of the subgraph $(V(G), X)$, with $c : 2^E \to \mathbb{R}_+$. 
Recall, $f : 2^V \rightarrow \mathbb{R}$ is submodular, then so is $\bar{f} : 2^V \rightarrow \mathbb{R}$ defined as $\bar{f}(S) = f(V \setminus S)$.

Hence, if $f : 2^V \rightarrow \mathbb{R}$ is supermodular, then so is $\bar{f} : 2^V \rightarrow \mathbb{R}$ defined as $\bar{f}(S) = f(V \setminus S)$.

Given a graph, for each $X \subseteq E(G)$, let $c(X)$ denote the number of connected components of the subgraph $(V(G), X)$, with $c : 2^E \rightarrow \mathbb{R}_+$. Then $c(X)$ is supermodular.
Other graph functions that are submodular/supermodular

- 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, for each $X \subseteq E(G)$, let $c(X)$ denote the number of connected components of the subgraph $(V(G), X)$, with $c : 2^E \to \mathbb{R}_+$. Then $c(X)$ is supermodular.

- $\bar{c}(X) = c(E \setminus X)$ is the number of connected components in $G$ when we remove $X$, and hence is also supermodular.
Graph Strength

- So $\bar{c}(X) = c(E \setminus X)$ is the number of connected components in $G$ when we remove $X$, is supermodular.
Graph Strength

- So $\overline{c}(X) = c(E \setminus X)$ is the number of connected components in $G$ when we remove $X$, is supermodular.
- Maximizing $\overline{c}(X)$ 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.
Graph Strength

- So $\overline{c}(X) = c(E \setminus X)$ is the number of connected components in $G$ when we remove $X$, is supermodular.
- Maximizing $\overline{c}(X)$ 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.
- If we can remove a small set $X$ and shatter the graph into many connected components, then the graph might be seen as being weak.
Graph Strength

- So $\bar{c}(X) = c(E \setminus X)$ is the number of connected components in $G$ when we remove $X$, is supermodular.
- Maximizing $\bar{c}(X)$ 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.
- If we can remove a small set $X$ and shatter the graph into many connected components, then the graph might be seen as being weak.
- An attacker should choose a small number of edges to shatter the graph into as many components as possible.
Graph Strength

- So $\bar{c}(X) = c(E \setminus X)$ is the number of connected components in $G$ when we remove $X$, is supermodular.
- Maximizing $\bar{c}(X)$ 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.
- If we can remove a small set $X$ and shatter the graph into many connected components, then the graph might be seen as being weak.
- An attacker should choose a small number of edges to shatter the graph into as many components as possible.
- Let $G = (V, E, w)$ with $w : E \to \mathbb{R}^+$ be a weighted graph with non-negative weights.
Graph Strength

- So $\bar{c}(X) = c(E \setminus X)$ is the number of connected components in $G$ when we remove $X$, is supermodular.

- Maximizing $\bar{c}(X)$ 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.

- If we can remove a small set $X$ and shatter the graph into many connected components, then the graph might be seen as being weak.

- An attacker should choose a small number of edges to shatter the graph into as many components as possible.

- Let $G = (V, E, w)$ with $w : E \to \mathbb{R}^+$ be a weighted graph with non-negative weights.

- For $(u, v) = e \in E$, let $w(e)$ be a measure of the strength of the connection between vertices $u$ and $v$ (strength meaning the difficulty of cutting the edge $e$).
Graph Strength

Then \( w(E) \) is a modular function

\[
\begin{align*}
  w(E) &= \sum_{e \in E} w_e \\
  \text{(2.34)}
\end{align*}
\]

so that \( w(E) \) is the “internal strength” of the vertex group \( S \).
Graph Strength

- Then \( w(E) \) is a modular function

\[
    w(E) = \sum_{e \in E} w_e
\]  \hspace{1cm} (2.34)

so that \( w(E) \) is the “internal strength” of the vertex group \( S \).

- Suppose removing \( E \) shatters \( G \) into a graph with \( \bar{c} > 1 \) components — then \( w(E)/(\bar{c} - 1) \) is like the “effort per achieved component” for a network attacker.
Graph Strength

- Then $w(E)$ is a modular function

$$w(E) = \sum_{e \in E} w_e \quad (2.34)$$

so that $w(E)$ is the “internal strength” of the vertex group $S$.

- Suppose removing $E$ shatters $G$ into a graph with $\bar{c} > 1$ components — then $w(E)/(\bar{c} - 1)$ is like the “effort per achieved component” for a network attacker.

- A form of graph strength can then be defined as the following:

$$\text{strength}(G, w) = \min_{A \subseteq E(G): \bar{c}(A) > 1} \frac{w(E(A))}{\bar{c}(A) - 1} \quad (2.35)$$
Graph Strength

- Then $w(E)$ is a modular function

$$w(E) = \sum_{e \in E} w_e$$

(2.34)

so that $w(E)$ is the “internal strength” of the vertex group $S$.

- Suppose removing $E$ shatters $G$ into a graph with $\bar{c} > 1$ components — then $w(E)/(\bar{c} - 1)$ is like the “effort per achieved component” for a network attacker.

- A form of graph strength can then be defined as the following:

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

(2.35)

- 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.
Let $V$ be an index set of a set of vectors in $\mathbb{R}^n$ for some $n$. 
Let $V$ be an index set of a set of vectors in $\mathbb{R}^n$ for some $n$.

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:

$$x_v = \sum_{i=1}^{k} \alpha_i x_{v_i} \quad (2.36)$$

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

- Let $V$ be an index set of a set of vectors in $\mathbb{R}^n$ for some $n$.
- 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:

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

(2.36)

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 $X = (x_1, x_2, \ldots, x_m)$. There are $m$ length-$n$ column vectors $\{x_i\}_i$. 

- 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 $X$. 

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

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

- Given $n \times m$ matrix $\mathbf{X} = (x_1, x_2, \ldots, x_m)$. 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 $\{x_a\}_{a \in A}$.

Skip matrix rank example
Example: Rank function of a matrix

- Given \( n \times m \) matrix \( \mathbf{X} = (x_1, x_2, \ldots, x_m) \). There are \( m \) length-\( n \) column vectors \( \{x_i\}_i \).
- Let \( V = \{1, 2, \ldots, m\} \) be the set of column vector indices.
- For any \( A \subseteq V \), let \( r(A) \) be the rank of the column vectors indexed by \( A \).
- \( r(A) \) is the dimensionality of the vector space spanned by the set of vectors \( \{x_a\}_{a \in A} \).
- Thus, \( r(V) \) is the rank of the matrix \( \mathbf{X} \).
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 $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$.

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

- 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 \(V = \{1, 2, 3, 4, 5, 6, 7, 8\}\).

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

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

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

\[
\begin{pmatrix}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
1 & 0 & 2 & 2 & 3 & 0 & 1 & 3 & 1 \\
2 & 0 & 3 & 0 & 4 & 0 & 0 & 2 & 4 \\
3 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 5 \\
4 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 5
\end{pmatrix}
\]

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

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

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

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

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

Prof. Jeff Bilmes
Example: Rank function of a matrix

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

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

- 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 $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$.

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

- 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 $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$.

\[
\begin{pmatrix}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
1 & 0 & 2 & 2 & 3 & 0 & 1 & 3 \\
2 & 0 & 3 & 0 & 4 & 0 & 0 & 2 \\
3 & 0 & 0 & 0 & 0 & 3 & 0 & 0 \\
4 & 2 & 0 & 0 & 0 & 0 & 0 & 5
\end{pmatrix}
= 
\begin{pmatrix}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
\mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 & \mathbf{x}_4 & \mathbf{x}_5 & \mathbf{x}_6 & \mathbf{x}_7 & \mathbf{x}_8
\end{pmatrix}
\]

- 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 $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$.

\[
\begin{array}{cccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
1 & 0 & 2 & 2 & 3 & 0 & 1 & 3 & 1 \\
2 & 0 & 3 & 0 & 4 & 0 & 0 & 2 & 4 \\
3 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 5 \\
4 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 5 \\
\end{array}
\]

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

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

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

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

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

\[
\begin{array}{cccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
1 & | & | & | & | & | & | & | \\
2 & | & | & | & | & | & | & | \\
3 & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 \\
4 & | & | & | & | & | & | & | \\
\end{array}
\]
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 $V = \{1, 2, 3, 4, 5, 6, 7, 8\}.$

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

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

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

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

- 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 $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$.

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

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

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

\[
\begin{bmatrix}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
1 & 0 & 2 & 2 & 3 & 1 & 3 & 1 \\
2 & 0 & 3 & 0 & 4 & 0 & 2 & 4 \\
3 & 0 & 0 & 0 & 3 & 0 & 0 & 5 \\
4 & 2 & 0 & 0 & 0 & 0 & 0 & 5
\end{bmatrix}
\]

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 $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$.

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

- 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.
- Let “area” correspond to dimensions spanned by vectors indexed by a set. Hence, $r(A)$ can be viewed as an area.

\[
\text{rank}(A) + \text{rank}(B) \geq \text{rank}(A \cup B)
\]
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.
- Let “area” correspond to dimensions spanned by vectors indexed by a set. Hence, \( r(A) \) can be viewed as an area.

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

\[
\begin{array}{c}
A \quad B \\
\text{Overlap} \\
A \cup B
\end{array}
\]
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.
- 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., $\exists$ common span), then that area 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.
- 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., $\exists$ common span), then that area 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.
Vectors $A$ and $B$ have a (possibly empty) common span and a (possibly empty) non-common residual span.
Rank functions of a matrix

- Vectors $A$ and $B$ have a (possibly empty) common span and a (possibly empty) non-common residual span.
- Let $C$ index vectors spanning dimensions common to $A$ and $B$. 

\begin{align*}
\text{Let } A_r \text{ index vectors spanning dimensions spanned by } A \text{ but not } B. \\
\text{Let } B_r \text{ index vectors spanning dimensions spanned by } B \text{ but not } A. \\
\text{Then, } r(A) = r(C) + r(A_r). \\
\text{Similarly, } r(B) = r(C) + r(B_r). \\
\text{Then } r(A) + r(B) \text{ counts the dimensions spanned by } C \text{ twice, i.e., } r(A) + r(B) = r(A_r) + 2r(C) + r(B_r). \\
\text{(2.37)} \\
\text{But } r(A \cup B) \text{ counts the dimensions spanned by } C \text{ only once. } \\
\text{r}(A \cup B) = r(A_r) + r(C) + r(B_r). \quad (2.38)
\end{align*}
Rank functions of a matrix

- Vectors $A$ and $B$ have a (possibly empty) common span and a (possibly empty) non-common residual span.
- Let $C$ index vectors spanning dimensions common to $A$ and $B$.
- Let $A_r$ index vectors spanning dimensions spanned by $A$ but not $B$. 

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

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

$r(A \cup B) = r(A_r) + r(C) + r(B_r)$
Vectors $A$ and $B$ have a (possibly empty) common span and a (possibly empty) non-common residual span.

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

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

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

- Vectors $A$ and $B$ have a (possibly empty) common span and a (possibly empty) non-common residual span.
- 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 a (possibly empty) non-common residual span.
- 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., $r(A) + r(B) = r(A_r) + 2r(C) + r(B_r)$. (2.37)

But $r(A \cup B)$ counts the dimensions spanned by $C$ only once. $r(A \cup B) = r(A_r) + r(C) + r(B_r)$. (2.38)
Rank functions of a matrix

- Vectors $A$ and $B$ have a (possibly empty) common span and a (possibly empty) non-common residual span.
- 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.,

$$r(A) + r(B) = r(A_r) + 2r(C) + r(B_r).$$  \hfill (2.37)
Rank functions of a matrix

- Vectors $A$ and $B$ have a (possibly empty) common span and a (possibly empty) non-common residual span.
- 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.,
  \[ r(A) + r(B) = r(A_r) + 2r(C) + r(B_r). \] (2.37)
- But $r(A \cup B)$ counts the dimensions spanned by $C$ only once.
  \[ r(A \cup B) = r(A_r) + r(C) + r(B_r) \] (2.38)
Rank functions of a matrix

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

\[
r(A) + r(B) = r(A_{r}) + 2r(C) + r(B_{r})
\]
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.
  \[
  r(A \cup B) = r(A_r) + r(C) + r(B_r)
  \]
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.
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 $C$).

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

In short:
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 $C$).

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

In short:
- Common span (blue) is no less than span of common index (magenta).
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 $C$).

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

In short:

- Common span (blue) is no less than span of common index (magenta).
- More generally, common information (blue) is no less than information within common index (magenta).
The Venn and Art of Submodularity

\[
\begin{align*}
 r(A) + r(B) & \geq r(A \cup B) + r(A \cap B) \\
& = r(A_r) + 2r(C) + r(B_r) \\
& = r(A_r) + r(C) + r(B_r) \\
& = r(A \cap B)
\end{align*}
\]
Let $S$ be a set of subspaces of a linear space and let, for each $X \subseteq S$, $f(X)$ denote the dimensionality of the linear subspace spanned by the subspaces in $X$. 

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 $A \subseteq B$).
Polymatroid rank function

- Let $S$ be a set of subspaces of a linear space and let, for each $X \subseteq S$, $f(X)$ 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 previous example, and for each $s \in S$, let $X_s$ being an index of vectors.
Polymatroid rank function

- Let \( S \) be a set of subspaces of a linear space and let, for each \( X \subseteq S \), \( f(X) \) 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 previous example, and for each \( s \in S \), let \( X_s \) being an index of vectors.
- Then, defining

\[
f(X) = r(\bigcup_{s \in S} X_s)
\]

(2.39)

is submodular, and is known to be a polymatroid rank function.
Polymatroid rank function

- Let $S$ be a set of subspaces of a linear space and let, for each $X \subseteq S$, $f(X)$ 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 previous example, and for each $s \in S$, let $X_s$ being an index of vectors.
- Then, defining

$$f(X) = r(\bigcup_{s \in S} X_s) \quad (2.39)$$

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 edges adjacent to $S$. 
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 edges adjacent to $S$.
- Then $r(S)$ is submodular, and is another matrix rank function corresponding to the incidence matrix of the graph.
Supply Side Economies of scale: Cost of manufacturing a set of items

- Let $V$ be a set of possible items that a company might possibly wish to manufacture, and let $f(S)$ for $S \subseteq V$ be the cost to that company to manufacture subset $S$.

- Ex: $V$ might be colors of paint in a paint manufacturer: green, red, blue, yellow, white, etc.

- Producing green when you are already producing yellow and blue is probably cheaper than if you were only producing some other colors.

  $$f(\text{green, blue, yellow}) - f(\text{blue, yellow}) \leq f(\text{green, blue}) - f(\text{blue})$$  

  \hspace{1cm} (2.5)

- So a submodular function would be a good model.
A model of Influence in Social Networks

- Given a graph $G = (V, E)$, each $v \in V$ corresponds to a person, to each $v$ we have an activation function $f_v : 2^V \rightarrow [0, 1]$ dependent only on its neighbors. I.e., $f_v(A) = f_v(A \cap \Gamma(v))$.

- Goal - Viral Marketing: find a small subset $S \subseteq V$ of individuals to directly influence, and thus indirectly influence the greatest number of possible other individuals (via the social network $G$).

- We define a function $f : 2^V \rightarrow \mathbb{Z}^+$ that models the ultimate influence of an initial set $S$ of nodes based on the following iterative process: At each step, a given set of nodes $S$ are activated, and we activate new nodes $v \in V \setminus S$ if $f_v(S) \geq U[0, 1]$ (where $U[0, 1]$ is a uniform random number between 0 and 1).

- It can be shown that for many $f_v$ (including simple linear functions, and where $f_v$ is submodular itself) that $f$ is submodular.
Let $V$ be a group of individuals. How valuable to you is a given friend $v \in V$?

It depends on how many friends you have.

Given a group of friends $S \subseteq V$, can you valuate them with a function $f(S)$ an how?

Let $f(S)$ be the value of the set of friends $S$. Is submodular or supermodular a good model?
Let $V$ be a set of information containing elements ($V$ might say be either words, sentences, documents, web pages, or blogs, each $v \in V$ is one element, so $v$ might be a word, a sentence, a document, etc.). The total amount of information in $V$ is measure by a a function $f(V)$, and any given subset $S \subseteq V$ measures the amount of information in $S$, given by $f(S)$.

How informative is any given item $v$ in different sized contexts? Any such real-world information function would exhibit diminishing returns, i.e., the value of $v$ decreases when it is considered in a larger context.

So a submodular function would likely be a good model.
Submodular Polyhedra

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

\[
P_f = \{ x \in \mathbb{R}^E : x(S) \leq f(S), \forall S \subseteq E \} \tag{2.40}
\]

\[
P_f^+ = P_f \cap \{ x \in \mathbb{R}^E : x \geq 0 \} \tag{2.41}
\]

\[
B_f = P_f \cap \{ x \in \mathbb{R}^E : x(E) = f(E) \} \tag{2.42}
\]

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

\[
\tilde{f}(c) \triangleq \max \{ c^T x : x \in P_f \} \tag{2.43}
\]

- This can be solved using the greedy algorithm! Moreover, \( \tilde{f}(c) \) computed using greedy is convex if and only of \( f \) is submodular (we will go into this in some detail this quarter).
Submodular functions are functions defined on subsets of some finite set, called the **ground set**.

- It is common in the literature to use either \( E \) or \( V \) as the ground set.
Submodular functions are functions defined on subsets of some finite set, called the ground set.

- It is common in the literature to use either $E$ or $V$ as the ground set.
- We will follow this inconsistency in the literature and will inconsistently use either $E$ or $V$ as our ground set (hopefully not in the same equation, if so, please point this out).
What does $x \in \mathbb{R}^E$ mean?

$$\mathbb{R}^E = \{ x = (x_j \in \mathbb{R} : j \in E) \} \quad (2.44)$$

$$\mathbb{R}_+^E = \{ x = (x_j : j \in E) : x \geq 0 \} \quad (2.45)$$

Any vector $x \in \mathbb{R}^E$ can be treated as a normalized modular function, and vice versa. That is

$$x(A) = \sum_{a \in A} x_a \quad (2.46)$$

Note that $x$ is said to be normalized since $x(\emptyset) = 0$. 
Other Notation: indicator vectors of sets

Given an \( A \subseteq E \), define the vector \( 1_A \in \mathbb{R}^E_+ \) to be

\[
1_A(j) = \begin{cases} 
1 & \text{if } j \in A; \\
0 & \text{if } j \notin A 
\end{cases}
\]  

(2.47)

Sometimes this will be written as \( \chi_A \equiv 1_A \).
Other Notation: indicator vectors of sets

Given an $A \subseteq E$, define the vector $\mathbf{1}_A \in \mathbb{R}^E_+$ to be

$$
\mathbf{1}_A(j) = \begin{cases} 
1 & \text{if } j \in A; \\
0 & \text{if } j \notin A 
\end{cases}
$$

(2.47)

Sometimes this will be written as $\chi_A \equiv \mathbf{1}_A$.

Thus, given modular function $x \in \mathbb{R}^E$, we can write $x(A)$ in a variety of ways, i.e.,

$$
x(A) = x \cdot \mathbf{1}_A = \sum_{i \in A} x(i)
$$

(2.48)
When $A$ is a set and $k$ is a singleton (i.e., a single item), the union is properly written as $A \cup \{k\}$, but sometimes I will write just $A + k$. 
General notation: what does $S^T$ mean when $S$ and $T$ are arbitrary sets

- Let $S$ and $T$ be two arbitrary sets (either of which could be countable, or uncountable).
General notation: what does $S^T$ mean when $S$ and $T$ are arbitrary sets

- Let $S$ and $T$ be two arbitrary sets (either of which could be countable, or uncountable).
- We define the notation $S^T$ to be the set of all functions that map from $T$ to $S$. That is, if $f \in S^T$, then $f : T \rightarrow S$. 
Let $S$ and $T$ be two arbitrary sets (either of which could be countable, or uncountable).

We define the notation $S^T$ to be the set of all functions that map from $T$ to $S$. That is, if $f \in S^T$, then $f : T \to S$.

Hence, given a finite set $E$, $\mathbb{R}^E$ is the set of all functions that map from elements of $E$ to the reals $\mathbb{R}$, and such functions are identical to a vector in a vector space with axes labeled as elements of $E$ (i.e., if $m \in \mathbb{R}^E$, then for all $e \in E$, $m(e) \in \mathbb{R}$).
General notation: what does $S^T$ mean when $S$ and $T$ are arbitrary sets

- Let $S$ and $T$ be two arbitrary sets (either of which could be countable, or uncountable).
- We define the notation $S^T$ to be the set of all functions that map from $T$ to $S$. That is, if $f \in S^T$, then $f : T \rightarrow S$.
- Hence, given a finite set $E$, $\mathbb{R}^E$ is the set of all functions that map from elements of $E$ to the reals $\mathbb{R}$, and such functions are identical to a vector in a vector space with axes labeled as elements of $E$ (i.e., if $m \in \mathbb{R}^E$, then for all $e \in E$, $m(e) \in \mathbb{R}$).
- Similarly, $2^E$ is the set of all functions from $E$ to “two” — in this case, we really mean $2 \equiv \{0, 1\}$, so $2^E$ is shorthand for $\{0, 1\}^V$. 
General notation: what does $S^T$ mean when $S$ and $T$ are arbitrary sets

- Let $S$ and $T$ be two arbitrary sets (either of which could be countable, or uncountable).
- We define the notation $S^T$ to be the set of all functions that map from $T$ to $S$. That is, if $f \in S^T$, then $f : T \rightarrow S$.
- Hence, given a finite set $E$, $\mathbb{R}^E$ is the set of all functions that map from elements of $E$ to the reals $\mathbb{R}$, and such functions are identical to a vector in a vector space with axes labeled as elements of $E$ (i.e., if $m \in \mathbb{R}^E$, then for all $e \in E$, $m(e) \in \mathbb{R}$).
- Similarly, $2^E$ is the set of all functions from $E$ to “two” — in this case, we really mean $2 \equiv \{0, 1\}$, so $2^E$ is shorthand for $\{0, 1\}^E$ — hence, $2^E$ is the set of all functions that map from elements of $E$ to $\{0, 1\}$, equivalent to all binary vectors with elements indexed by elements of $E$, equivalent to subsets of $E$. Hence, if $A \in 2^E$ then $A \subseteq E$. 
Given $E$, let $f_1, f_2 : 2^E \to \mathbb{R}$ be two submodular functions. Then

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

(2.49)

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

$$f : 2^E \to \mathbb{R} \text{ with } f(A) = f_1(A) + f_2(A) \quad (2.49)$$

is submodular. This follows easily since

$$f(A) + f(B) = f_1(A) + f_2(A) + f_1(B) + f_2(B) \quad (2.50)$$

$$\geq f_1(A \cup B) + f_2(A \cup B) + f_1(A \cap B) + f_2(A \cap B) \quad (2.51)$$

$$= f(A \cup B) + f(A \cap B). \quad (2.52)$$

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

$$f : 2^E \rightarrow \mathbb{R} \text{ with } f(A) = f_1(A) + f_2(A) \quad (2.49)$$

is submodular. This follows easily since

$$f(A) + f(B) = f_1(A) + f_2(A) + f_1(B) + f_2(B) \quad (2.50)$$

$$\geq f_1(A \cup B) + f_2(A \cup B) + f_1(A \cap B) + f_2(A \cap B) \quad (2.51)$$

$$= f(A \cup B) + f(A \cap B). \quad (2.52)$$

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 $\alpha_1, \alpha_2 \geq 0$. 
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

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

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

\[ f : 2^E \rightarrow \mathbb{R} \text{ with } f(A) = f_1(A) - m(A) \quad (2.53) \]

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

\[
\begin{align*}
  f(A) + f(B) &= f_1(A) - m(A) + f_1(B) - m(B) \\
  &\geq f_1(A \cup B) - m(A \cup B) + f_1(A \cap B) - m(A \cap B) \\
  &= f(A \cup B) + f(A \cap B).
\end{align*}
\]
Summing Submodular and Modular Functions

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

$$f : 2^E \rightarrow \mathbb{R} \text{ with } f(A) = f_1(A) - m(A)$$  \hspace{1cm} (2.53)

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

$$f(A) + f(B) = f_1(A) - m(A) + f_1(B) - m(B)$$
$$\geq f_1(A \cup B) - m(A \cup B) + f_1(A \cap B) - m(A \cap B)$$ \hspace{1cm} (2.55)
$$= f(A \cup B) + f(A \cap B).$$ \hspace{1cm} (2.56)

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

\[
f' : 2^E \rightarrow \mathbb{R} \text{ with } f'(A) = f(A \cap S) \tag{2.57}
\]

is submodular.
Restricting Submodular Functions

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

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

is submodular.

Proof.
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

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

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

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

is submodular.

Proof.

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

$$f((A + v) \cap S) - f(A \cap S) \geq f((B + v) \cap S) - f(B \cap S) \tag{2.58}$$

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

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

is submodular.

**Proof.**

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

$$f((A + v) \cap S) - f(A \cap S) \geq f((B + v) \cap S) - f(B \cap S) \tag{2.58}$$

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

$$f(A' + v) - f(A') \geq f(B' + v) - f(B') \tag{2.59}$$

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

Summing Restricted Submodular Functions

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

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

(2.60)

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

$$f : 2^V \rightarrow \mathbb{R} \text{ with } f(A) = f_1(A \cap S_1) + f_2(A \cap S_2) \quad (2.60)$$

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

$$f : 2^V \rightarrow \mathbb{R} \text{ with } f(A) = \sum_{C \in C} f_C(A \cap C) \quad (2.61)$$

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

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

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

(2.62)

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

Proof.

Consider

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

(2.63)

which follows since we have that

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

(2.64)

and

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

(2.65)
Given $V$, let $c \in \mathbb{R}^V$ be a given fixed vector (not necessarily non-negative). Then $f : 2^V \to \mathbb{R}$, where

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

(2.66)

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)

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 needing to be serviced (e.g., cities, clients).

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 needs to 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 \( f(A) = \sum_{j \in A} m_j + \sum_{i \in F} \max_{j \in A} c_{ij} \).

Goal is to find a set \( A \) that maximizes \( f(A) \) (the benefit) placing a bound on the number of plants \( A \) (e.g., \( |A| \leq k \)).
Given $V, E$, let $c \in \mathbb{R}^{V \times E}$ be a given $|V| \times |E|$ matrix. Then

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

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 $A$.

$$f(A) = \log \det(\Sigma_A) \text{ is submodular.} \quad (2.68)$$

**Proof.**

Suppose $x \in \mathbb{R}^n$ is multivariate Gaussian, that is

$$x \in p(x) = \frac{1}{\sqrt{|2\pi\Sigma|}} \exp \left( -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right) \quad (2.69)$$

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

$$h(X) = \log \sqrt{|2\pi e \Sigma|} = \log \sqrt{(2\pi e)^n |\Sigma|}$$  \hspace{1cm} (2.70)

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

$$f(A) = h(X_A) = \log \sqrt{(2\pi e)^{|A|} |\Sigma_A|}$$  \hspace{1cm} (2.71)

Entropy is submodular (conditioning reduces entropy), and moreover

$$f(A) = h(X_A) = m(A) + \frac{1}{2} \log |\Sigma_A|$$  \hspace{1cm} (2.72)

where $m(A)$ is a modular function.

Note: still submodular in the semi-definite case as well.
Concave over non-negative modular

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

\[
f(A) = g(m(A)) \tag{2.73}
\]

then \( f \) is submodular.

**Proof.**

Given \( A \subseteq B \subseteq E \setminus v \), we have \( 0 \leq a = m(A) \leq b = m(B) \), and

\( 0 \leq c = m(v) \). For \( g \) concave, we have \( g(a + c) - g(a) \geq g(b + c) - g(b) \), and thus

\[
g(m(A) + m(v)) - g(m(A)) \geq g(m(B) + m(v)) - g(m(B)) \tag{2.74}
\]

A form of converse is true as well.