Submodular Functions, Optimization, and Applications to Machine Learning

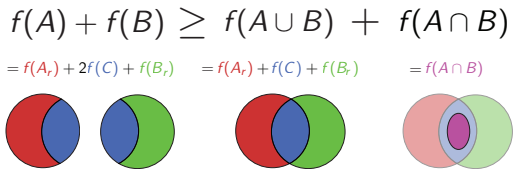
— Fall Quarter, Lecture 4 —

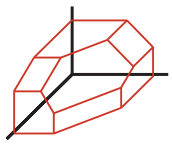
http://www.ee.washington.edu/people/faculty/bilmes/classes/ee563_spring_2018/

Prof. Jeff Bilmes

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

Oct 12th, 2020





Announcements, Assignments, and Reminders

- Homework 1 is out, due Friday at 11:59pm.

Class Road Map - EE563

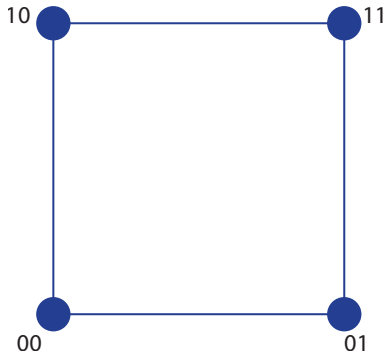
- L1(9/30): Motivation, Applications, Definitions, Properties
- L2(10/5): Sums concave(modular), uses (diversity/costs, feature selection), information theory
- L3(10/7): Monge, More Definitions, Graph and Combinatorial Examples,
- L4(10/12): Graph & Combinatorial Examples, Matrix Rank, Properties, Other Defs, Independence
- L5(10/14):
- L6(10/19):
- L7(10/21):
- L8(10/26):
- L9(10/28):
- L10(11/2):
- L11(11/4):
- L12(11/9):
- L–(11/11): Veterans Day, Holiday
- L13(11/16):
- L14(11/18):
- L15(11/23):
- L16(11/25):
- L17(11/30):
- L18(12/2):
- L19(12/7):
- L20(12/9): maximization.

Last day of instruction, Fri. Dec 11th. Finals Week: Dec 12-18, 2020

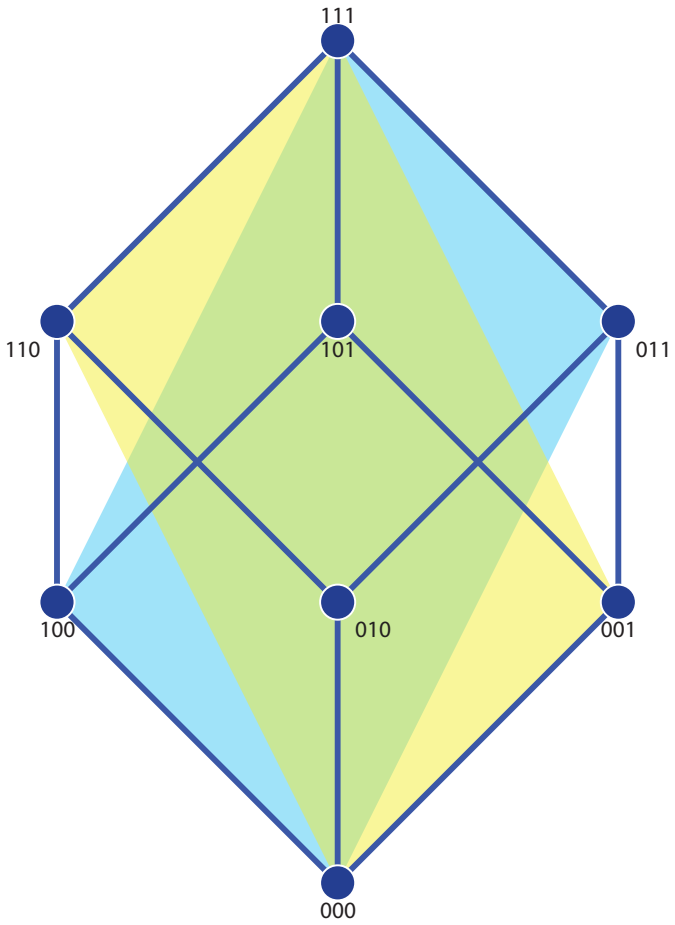
The Submodular Square, and Hypercube Vertices

We can test submodularity via values on vertices of hypercube.

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



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



How many inequalities of form $f(A) + f(B) \geq f(A \cup B) + f(A \cap B)$?

Subadditive Definitions

Definition 4.2.1 (subadditive)

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

$$f(A) + f(B) \geq f(A \cup B) \tag{4.7}$$

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

Superadditive Definitions

Definition 4.2.1 (superadditive)

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

$$f(A) + f(B) \leq f(A \cup B) \tag{4.7}$$

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

$$f(A) = \begin{cases} 1 & \text{if } |A| \leq k \\ 0 & \text{else} \end{cases} \tag{4.8}$$

- This function is subadditive but not submodular.

Modular Definitions

Definition 4.2.1 (modular)

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

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

$$f(A) + f(B) = f(A \cap B) + f(A \cup B) \tag{4.7}$$

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

Proposition 4.2.2

If f is modular, it may be written as

$$f(A) = f(\emptyset) + \sum_{a \in A} \Big(f(\{a\}) - f(\emptyset) \Big) = c + \sum_{a \in A} f'(a) \tag{4.8}$$

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

Complement function

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

Proposition 4.2.1

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

Proof.

$$\bar{f}(A) + \bar{f}(B) \geq \bar{f}(A \cup B) + \bar{f}(A \cap B) \tag{4.12}$$

follows from

$$f(V \setminus A) + f(V \setminus B) \geq f(V \setminus (A \cup B)) + f(V \setminus (A \cap B)) \tag{4.13}$$

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 number of edges with exactly one vertex in $S \subseteq V(G)$ is submodular (cut function). Thus, we have $I(S) = E(S) \cup \delta(S)$ and $E(S) \cap \delta(S) = \emptyset$, and thus that $|I(S)| = |E(S)| + |\delta(S)|$. So we can get a submodular function by summing a submodular and a supermodular function. If you had to guess, is this always the case?
- Consider $f(A) = |\delta^+(A)| - |\delta^+(V \setminus A)|$. Guess, submodular, supermodular, modular, or neither? Exercise: determine which one and prove it.

Number of connected components in a graph via edges

- Recall, $f : 2^V \to \mathbb{R}$ is submodular, then so is $\bar{f} : 2^V \to \mathbb{R}$ defined as $\bar{f}(S) = f(V \setminus S)$.
- Hence, if $g : 2^V \to \mathbb{R}$ is supermodular, then so is $\bar{g} : 2^V \to \mathbb{R}$ defined as $\bar{g}(S) = g(V \setminus S)$.
- Given a graph $G = (V, E)$, for each $A \subseteq E(G)$, let $c(A)$ denote the number of connected components of the (spanning) subgraph $(V(G), A)$, with $c : 2^E \to \mathbb{R}_+$. Thus, $c(\emptyset) = |V|$, and $c(E) \geq 1$.
- $c(A)$ is monotone non-increasing, $c(A + a) - c(A) \leq 0$.
- Then $c(A)$ is supermodular, i.e.,
$$c(A + a) - c(A) \leq c(B + a) - c(B) \tag{4.26}$$
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 number of connected components in G when we remove A; supermodular monotone non-decreasing but not normalized.

Graph Strength

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

$$w(A) = \sum_{e \in A} w_e \tag{4.26}$$

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

$$\textit{strength}(G, w) \triangleq \min_{A \subseteq E(G): \bar{c}(A) > 1} \frac{w(A)}{\bar{c}(A) - 1} \tag{4.27}$$

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

Quadratic forms

- Consider $\phi : \mathbb{R}^n \to R$ with

$$\phi(x) = \langle m, x \rangle + \frac{1}{2} x^\intercal M x \tag{4.1}$$

M is an nxn matrix

Quadratic forms

- Consider $\phi : \mathbb{R}^n \to R$ with

$$\phi(x) = \langle m, x \rangle + \frac{1}{2} x^\intercal M x \tag{4.1}$$

- $\phi(x)$ is convex iff M is positive semidefinite (requiring all diagonal entries to be non-negative).

Quadratic forms

- Consider $\phi : \mathbb{R}^n \to R$ with

$$\phi(x) = \langle m, x \rangle + \frac{1}{2} x^\intercal M x \tag{4.1}$$

- $\phi(x)$ is convex iff M is positive semidefinite (requiring all diagonal entries to be non-negative).
- $\phi(x)$ is concave iff M is negative semidefinite (requiring all diagonal entries to be non-positive)

Quadratic forms

- Consider $\phi : \mathbb{R}^n \to R$ with

$$\phi(x) = \langle m, x \rangle + \frac{1}{2} x^\intercal M x \tag{4.1}$$

- $\phi(x)$ is convex iff M is symmetric positive semidefinite (requiring all diagonal entries to be non-negative).
- $\phi(x)$ is concave iff M is symmetric negative semidefinite (requiring all diagonal entries to be non-positive)
- A matrix M with the requirement only of having non-positive off-diagonal entries could lead to a function that is either convex, concave, or neither. Exercise.

Submodularity, Quadratic Structures, and Cuts

Lemma 4.3.1

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

$$f(X) = \overbrace{m^\intercal \mathbf{1}_X}^{m(X)} + \frac{1}{2}\mathbf{1}_X^\intercal \mathbf{M}\mathbf{1}_X \tag{4.2}$$

is submodular <u>*iff*</u> *the off-diagonal elements of* M *are non-positive.*

Proof.

1) diff. requirement on m then concave or convex.

2) latter, see how this is related to lattice theory & DR submodularity.

...

Submodularity, Quadratic Structures, and Cuts

Lemma 4.3.1

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

$$f(X) = m^\intercal \mathbf{1}_X + \frac{1}{2}\mathbf{1}_X^\intercal \mathbf{M} \mathbf{1}_X \tag{4.2}$$

is submodular <u>*iff*</u> *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.

$$|E(X)| = \sum_{i,j \in X} 1$$

...

Submodularity, Quadratic Structures, and Cuts

Lemma 4.3.1

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

$$f(X) = m^\intercal \mathbf{1}_X + \frac{1}{2}\mathbf{1}_X^\intercal \mathbf{M} \mathbf{1}_X \tag{4.2}$$

is submodular <u>*iff*</u> *the off-diagonal elements of* M *are non-positive.*

Proof.

- Given a complete graph $G = (V, E)$, recall that $E(X)$ is the edge set with both vertices in $X \subseteq V(G)$, and that $|E(X)|$ is supermodular.
- Non-negative modular weights $w^+ : E \to \mathbb{R}_+$, $w(E(X))$ is also supermodular, so $-w(E(X))$ is submodular.

...

Submodularity, Quadratic Structures, and Cuts

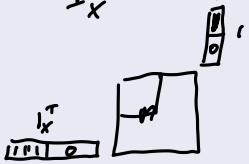
Lemma 4.3.1

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

$1_X^T \cdot M \cdot 1_X = \sum_{i,j \in X} m_{ij}$

$$f(X) = m^\intercal \mathbf{1}_X + \frac{1}{2}\mathbf{1}_X^\intercal \mathbf{M} \mathbf{1}_X \tag{4.2}$$

$\| \quad \sum_{i,j \in X} m_{ij} = \sum_{\substack{i,j \in X \\ i \neq j}} m_{ij} + \sum_{i \in X} m_{ii}$



is submodular <u>*iff*</u> *the off-diagonal elements of* M *are non-positive.*

Proof.

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

...

Submodularity, Quadratic Structures, and Cuts

Proof of Lemma 4.3.1 cont.

- Conversely, suppose f is submodular. Consider $u, v \in V$ with $u \neq v$.

$\square$

Submodularity, Quadratic Structures, and Cuts

Proof of Lemma 4.3.1 cont.

- Conversely, suppose f is submodular. Consider $u, v \in V$ with $u \neq v$.
- Then $f(\{u\}) + f(\{v\}) \geq f(\{u, v\}) + f(\emptyset)$ and $f(\emptyset) = 0$

$\square$

Submodularity, Quadratic Structures, and Cuts

Proof of Lemma 4.3.1 cont.

- Conversely, suppose f is submodular. Consider $u, v \in V$ with $u \neq v$.
- Then $f(\{u\}) + f(\{v\}) \geq f(\{u, v\}) + f(\emptyset)$ and $f(\emptyset) = 0$ (note that $u = v$ would lead to a vacuous statement).

$\square$

Submodularity, Quadratic Structures, and Cuts

Proof of Lemma 4.3.1 cont.

- Conversely, suppose f is submodular. Consider $u, v \in V$ with $u \neq v$.
- Then $f(\{u\}) + f(\{v\}) \geq f(\{u, v\}) + f(\emptyset)$ and $f(\emptyset) = 0$ (note that $u = v$ would lead to a vacuous statement).
- This requires:

$$
\begin{aligned}
0 &\leq f(\{u\}) + f(\{v\}) - f(\{u, v\}) && (4.3)\\
&= m(u) + \frac{1}{2}M_{u,u} + m(v) + \frac{1}{2}M_{v,v} && (4.4)\\
&\quad - \left(m(u) + m(v) + \frac{1}{2}M_{u,u} + M_{u,v} + \frac{1}{2}M_{v,v}\right) && (4.5)\\
&= -M_{u,v} && (4.6)
\end{aligned}
$$

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

□

Set Cover and Maximum Coverage

just Special cases of Submodular Optimization

- We are given a finite set U of m elements and a set of subsets $\mathcal{U} = \{U_1, U_2, \dots, U_n\}$ of n subsets of U, so that $U_i \subseteq U$ and $\bigcup_i U_i = U$.

Set Cover and Maximum Coverage

just Special cases of Submodular Optimization

- We are given a finite set U of m elements and a set of subsets $\mathcal{U} = \{U_1, U_2, \dots, U_n\}$ of n subsets of U, so that $U_i \subseteq U$ and $\bigcup_i U_i = U$.
- The goal of minimum set cover is to choose the smallest subset $A \subseteq [n] \triangleq \{1, \dots, n\}$ such that $\bigcup_{a \in A} U_a = U$.

Set Cover and Maximum Coverage

just Special cases of Submodular Optimization

- We are given a finite set U of m elements and a set of subsets $\mathcal{U} = \{U_1, U_2, \dots, U_n\}$ of n subsets of U, so that $U_i \subseteq U$ and $\bigcup_i U_i = U$.
- The goal of minimum set cover is to choose the smallest subset $A \subseteq [n] \triangleq \{1, \dots, n\}$ such that $\bigcup_{a \in A} U_a = U$.
- Maximum k cover: The goal in maximum coverage is, given an integer $k \leq n$, select k subsets, say $\{a_1, a_2, \dots, a_k\}$ with $a_i \in [n]$ such that $|\bigcup_{i=1}^{k} U_{a_i}|$ is maximized.

Set Cover and Maximum Coverage

just Special cases of Submodular Optimization

- We are given a finite set U of m elements and a set of subsets $\mathcal{U} = \{U_1, U_2, \dots, U_n\}$ of n subsets of U, so that $U_i \subseteq U$ and $\bigcup_i U_i = U$.
- The goal of minimum set cover is to choose the smallest subset $A \subseteq [n] \triangleq \{1, \dots, n\}$ such that $\bigcup_{a \in A} U_a = U$.
- Maximum k cover: The goal in maximum coverage is, given an integer $k \leq n$, select k subsets, say $\{a_1, a_2, \dots, a_k\}$ with $a_i \in [n]$ such that $|\bigcup_{i=1}^k U_{a_i}|$ is maximized.
- $f : 2^{[n]} \to \mathbb{Z}_+$ where for $A \subseteq [n]$, $f(A) = |\bigcup_{a \in A} U_a|$ is the set cover function and is submodular.

Set Cover and Maximum Coverage

just Special cases of Submodular Optimization

$[n] = \{1, 2, 3, \ldots, n\}$

- We are given a finite set U of m elements and a set of subsets $\mathcal{U} = \{U_1, U_2, \ldots, U_n\}$ of n subsets of U, so that $U_i \subseteq U$ and $\bigcup_i U_i = U$.
- The goal of minimum set cover is to choose the smallest subset $A \subseteq [n] \triangleq \{1, \ldots, n\}$ such that $\bigcup_{a \in A} U_a = U$.
- Maximum k cover: The goal in maximum coverage is, given an integer $k \leq n$, select k subsets, say $\{a_1, a_2, \ldots, a_k\}$ with $a_i \in [n]$ such that $|\bigcup_{i=1}^{k} U_{a_i}|$ is maximized.
- $f : 2^{[n]} \to \mathbb{Z}_+$ where for $A \subseteq [n]$, $f(A) = |\bigcup_{a \in A} U_a|$ is the set cover function and is submodular.
- Weighted set cover: $f(A) = w(\bigcup_{a \in A} U_a)$ where $w : U \to \mathbb{R}_+$.

Set Cover and Maximum Coverage

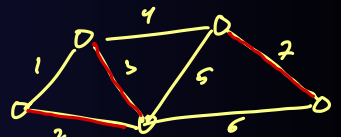
just Special cases of Submodular Optimization

- We are given a finite set U of m elements and a set of subsets $\mathcal{U} = \{U_1, U_2, \dots, U_n\}$ of n subsets of U, so that $U_i \subseteq U$ and $\bigcup_i U_i = U$.
- The goal of minimum set cover is to choose the smallest subset $A \subseteq [n] \triangleq \{1, \dots, n\}$ such that $\bigcup_{a \in A} U_a = U$.
- Maximum k cover: The goal in maximum coverage is, given an integer $k \leq n$, select k subsets, say $\{a_1, a_2, \dots, a_k\}$ with $a_i \in [n]$ such that $|\bigcup_{i=1}^k U_{a_i}|$ is maximized.
- $f : 2^{[n]} \to \mathbb{Z}_+$ where for $A \subseteq [n]$, $f(A) = |\bigcup_{a \in A} U_a|$ is the set cover function and is submodular.
- Weighted set cover: $f(A) = w(\bigcup_{a \in A} U_a)$ where $w : U \to \mathbb{R}_+$.
- Both Set cover and maximum coverage are well known to be NP-hard, but have a fast greedy approximation algorithm, and hence are instances of submodular optimization.

Vertex and Edge Covers

Also instances of submodular optimization

$S = \{a, c, d\}$ is a vertex cover.



$A = \{4, 3, 7\}$ is an edge cover.

plastic cover, the plastic is doing the covering.

Definition 4.3.2 (vertex cover), the vertices are doing the covering.

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), the edge are doing the covering

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.

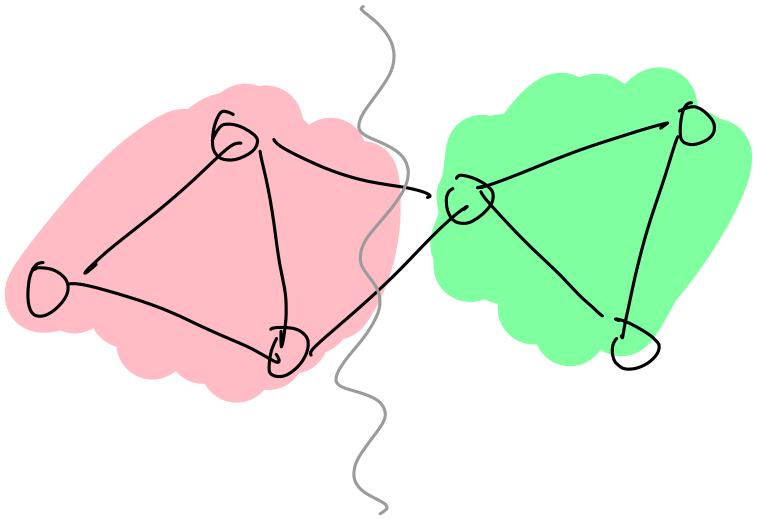
$|V(F)|$

- 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 maximize 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 maximize the cut (set of edges) between S and $V \setminus S$.
- Let $\delta : 2^V \to \mathbb{R}_+$ be the cut function, namely for any given set of nodes $X \subseteq V$, $|\delta(X)|$ measures the number of edges between nodes X and $V \setminus X$ — i.e., $\delta(X) = E(X, V \setminus X)$.

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 maximize the cut (set of edges) between S and $V \setminus S$.
- Let $\delta : 2^V \to \mathbb{R}_+$ be the cut function, namely for any given set of nodes $X \subseteq V$, $|\delta(X)|$ measures the number of edges between nodes X and $V \setminus X$ — i.e., $\delta(x) = E(X, V \setminus X)$.
- Weighted versions, where rather than count, we sum the (non-negative) weights of the edges of a cut, $f(X) = w(\delta(X))$.

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 maximize the cut (set of edges) between S and $V \setminus S$.
- Let $\delta : 2^V \to \mathbb{R}_+$ be the cut function, namely for any given set of nodes $X \subseteq V$, $|\delta(X)|$ measures the number of edges between nodes X and $V \setminus X$ — i.e., $\delta(x) = E(X, V \setminus X)$.
- Weighted versions, where rather than count, we sum the (non-negative) weights of the edges of a cut, $f(X) = w(\delta(X))$.
- Hence, Minimum cut and Maximum cut are also special cases of submodular optimization.

Matrix Rank functions

- Let V, with $|V| = n$ be an index set of a set of vectors in $\mathbb{R}^m$ for some m (unrelated to n). Thus, $\forall v \in V$, $\exists x_v \in \mathbb{R}^m$.

Matrix Rank functions

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

x_{v_i}

$$x_v = \sum_{i=1}^{k} \alpha_i x_{v_i} \tag{4.7}$$

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

Matrix Rank functions

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

$$x_v = \sum_{i=1}^{k} \alpha_i x_{v_i} \tag{4.7}$$

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

Example: Rank function of a matrix

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

$$X = \begin{bmatrix} | & | & & | \\ x_1 & x_2 & \cdots & x_n \\ | & | & & | \end{bmatrix} \quad m$$

Skip matrix rank example

Example: Rank function of a matrix

- Given $m \times n$ matrix $\mathbf{X} = (x_1, x_2, \ldots, x_n)$ with $x_i \in \mathbb{R}^m$ for all i. There are n length-m column vectors $\{x_i\}_i$
- Let $V = \{1, 2, \ldots, n\}$ be the set of column vector indices.

Skip matrix rank example

Example: Rank function of a matrix

- Given $m \times n$ matrix $\mathbf{X} = (x_1, x_2, \ldots, x_n)$ with $x_i \in \mathbb{R}^m$ for all i. There are n length-m column vectors $\{x_i\}_i$
- Let $V = \{1, 2, \ldots, n\}$ 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.

Skip matrix rank example

Example: Rank function of a matrix

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

Skip matrix rank example

Example: Rank function of a matrix

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

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

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

Example: Rank function of a matrix

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

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

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

Example: Rank function of a matrix

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

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

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

Example: Rank function of a matrix

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

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

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

Example: Rank function of a matrix

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

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

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

Example: Rank function of a matrix

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

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

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

Example: Rank function of a matrix

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

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

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

Example: Rank function of a matrix

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

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

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

Example: Rank function of a matrix

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

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

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

Example: Rank function of a matrix

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

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

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

Example: Rank function of a matrix

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

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

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

  C ~~is~~ spans the common spanned dimension of A & B.
- $r(A \cup A_r) = 3$, $r(B \cup B_r) = 3$, $r(A \cup B_r) = 4$, $r(B \cup A_r) = 4$.
- $r(A \cup B) = 4$, $r(A \cap B) = 1 < r(C) = 2$.

Example: Rank function of a matrix

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

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

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

Example: Rank function of a matrix

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

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

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

Example: Rank function of a matrix

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

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

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

Example: Rank function of a matrix

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

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

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

Example: Rank function of a matrix

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

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

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

Example: Rank function of a matrix

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

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

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

common span

Example: Rank function of a matrix

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

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

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

Example: Rank function of a matrix

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

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

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

Rank function of a matrix

- Let $A, B \subseteq V$ be two subsets of column indices.

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.

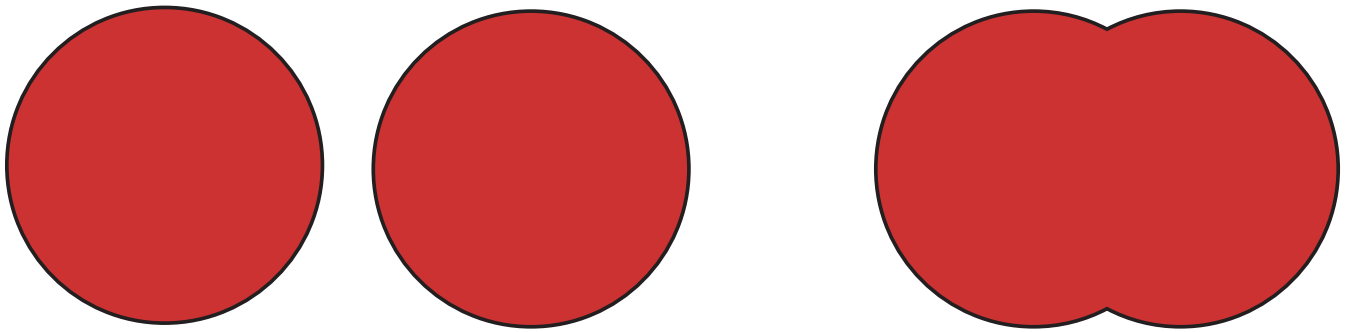
Rank function of a matrix

- 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.

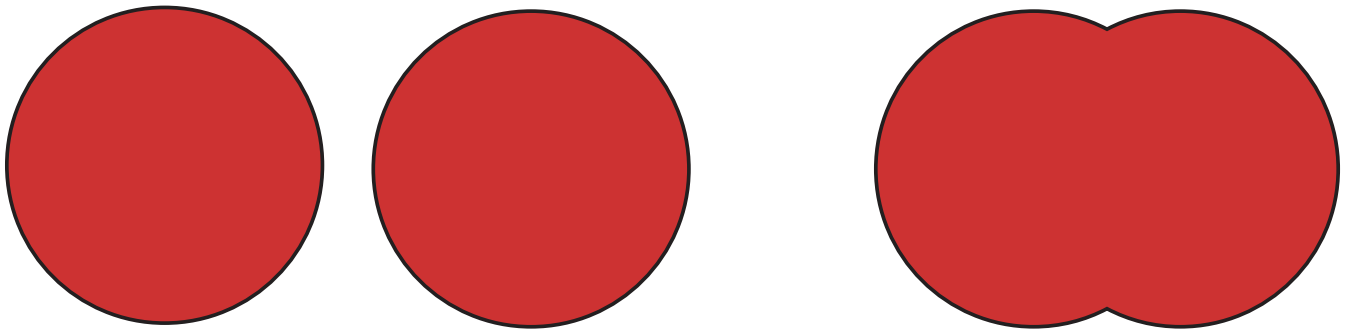
$$r(A) \quad + \quad r(B) \qquad \geq \qquad r(A \cup B)$$



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.

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

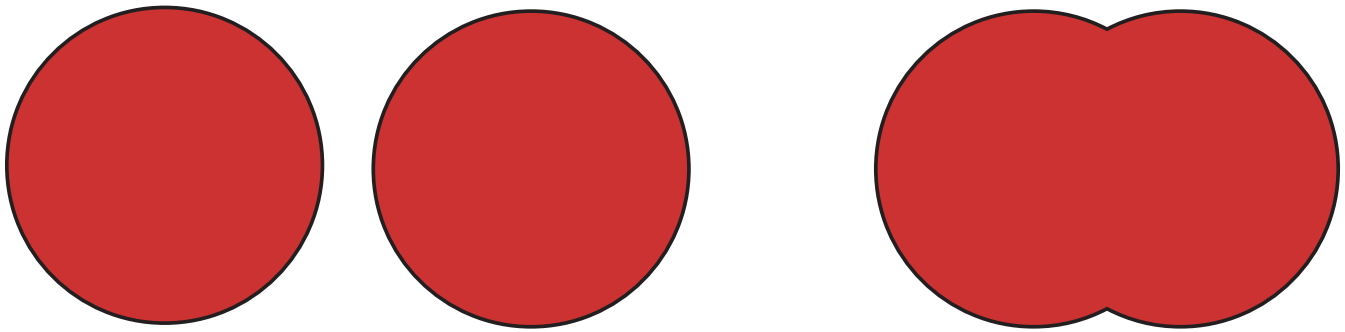


- If some of the dimensions spanned by A overlap some of the dimensions spanned by B (i.e., if $\exists$ common span), then that area is counted twice in $r(A) + r(B)$, so the inequality will be strict.

Rank function of a matrix

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

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



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

Rank functions of a matrix

- Vector sets A and B have a (possibly empty) common span and two (possibly empty) non-common residual spans.

Rank functions of a matrix

- Vector sets A and B have a (possibly empty) common span and two (possibly empty) non-common residual spans.
- Let C index vectors spanning all dimensions common to A and B. We call C the common span and call $A \cap B$ the common indices.

Rank functions of a matrix

- Vector sets A and B have a (possibly empty) common span and two (possibly empty) non-common residual spans.
- Let C index vectors spanning all dimensions common to A and B. We call C the common span and call $A \cap B$ the common indices.
- Let A_r index vectors spanning dimensions spanned by A but not B.

Rank functions of a matrix

- Vector sets A and B have a (possibly empty) common span and two (possibly empty) non-common residual spans.
- Let C index vectors spanning all dimensions common to A and B. We call C the common span and call $A \cap B$ the common indices.
- 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

- Vector sets A and B have a (possibly empty) common span and two (possibly empty) non-common residual spans.
- Let C index vectors spanning all dimensions common to A and B. We call C the common span and call $A \cap B$ the common indices.
- 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

- Vector sets A and B have a (possibly empty) common span and two (possibly empty) non-common residual spans.
- Let C index vectors spanning all dimensions common to A and B. We call C the common span and call $A \cap B$ the common indices.
- 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, $r(B) = r(C) + r(B_r)$.

Rank functions of a matrix

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

$$r(A) + r(B) = r(A_r) + 2r(C) + r(B_r). \tag{4.8}$$

Rank functions of a matrix

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

$$r(A) + r(B) = r(A_r) + 2r(C) + r(B_r). \tag{4.8}$$

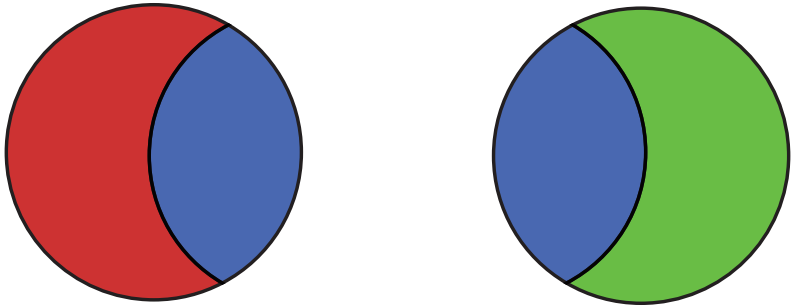
- But $r(A \cup B)$ counts the dimensions spanned by C only once.

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

Rank functions of a matrix

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

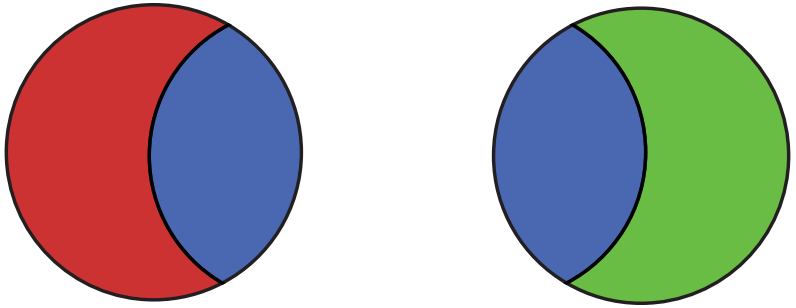
$$r(A) + r(B) = r(A_r) + 2r(C) + r(B_r)$$



Rank functions of a matrix

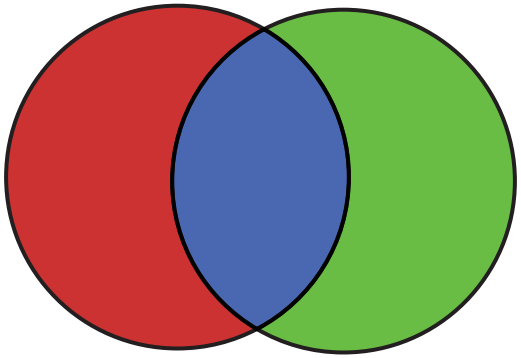
- Then $r(A) + r(B)$ counts the dimensions spanned by C twice, i.e.,

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



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

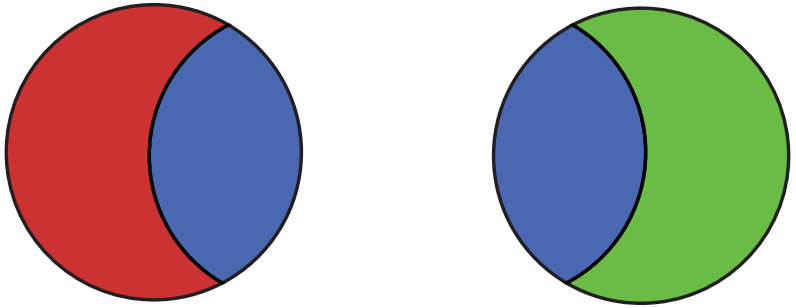
$$r(A \cup B) = r(A_r) + r(C) + r(B_r)$$



Rank functions of a matrix

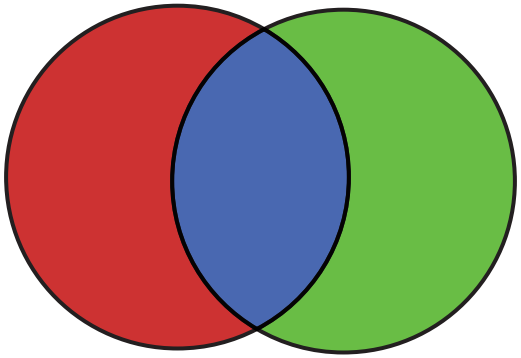
- Then $r(A) + r(B)$ counts the dimensions spanned by C twice, i.e.,

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



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

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

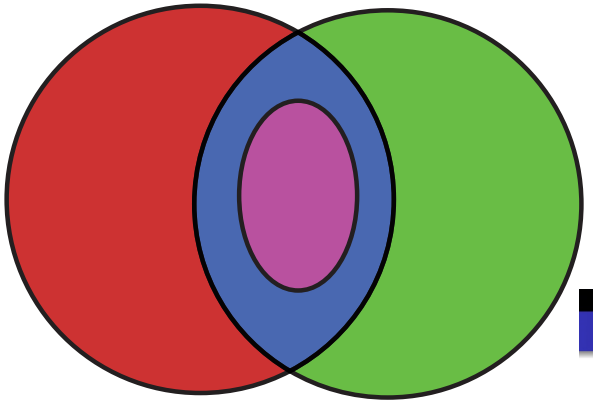


- Thus, we have subadditivity: $r(A) + r(B) \geq r(A \cup B)$. Can we add more to the r.h.s. and still have an inequality? Yes.

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:

Example: Rank function of a matrix

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

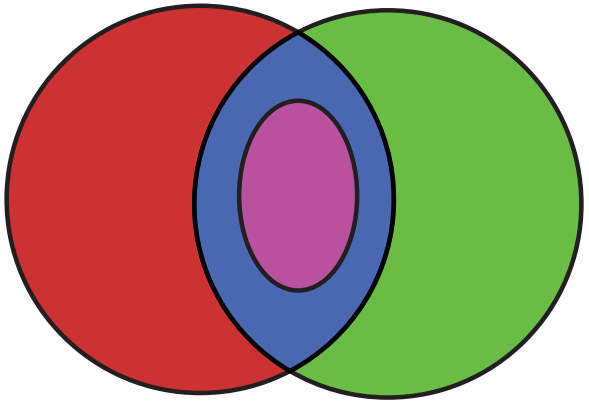
$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{pmatrix} 0 & 2 & 2 & 3 & 0 & 1 & 3 & 1 \\ 0 & 3 & 0 & 4 & 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 & 3 & 0 & 0 & 5 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 5 \end{pmatrix} \end{array}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

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

Rank function of a matrix

- 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).

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



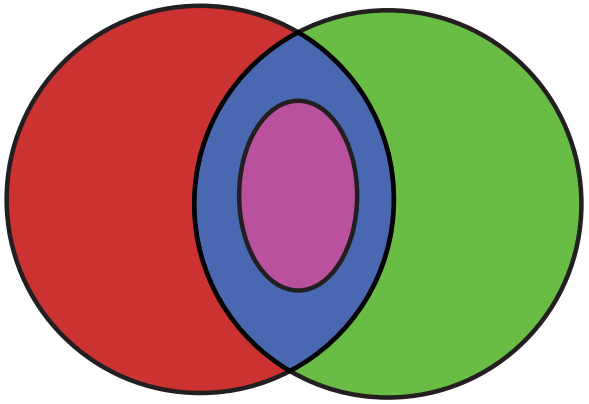
In short:

- 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).

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



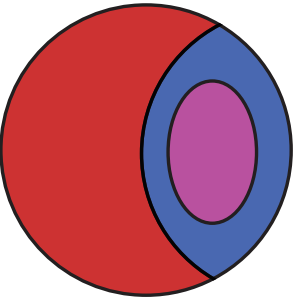
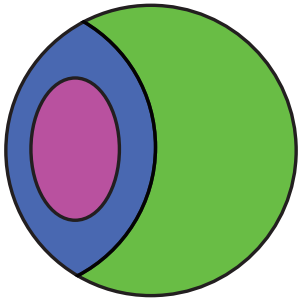
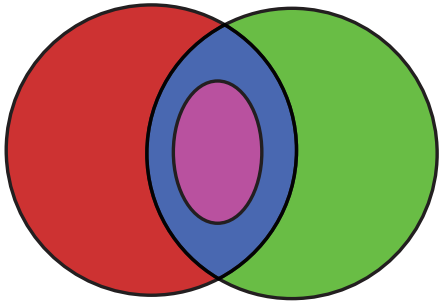
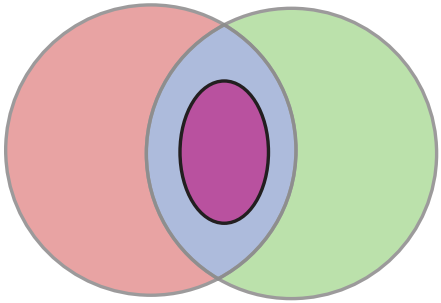
In short:

- 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




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

Polymatroid rank function

- Let S be a set of subspaces of a linear space (i.e., each $s \in S$ is a subspace of dimension ≥ 1).

Polymatroid rank function

- Let S be ~~a set of subspaces~~ an index set of a of a linear space (i.e., each $s \in S$ is a subspace of dimension ≥ 1).
- For each $X \subseteq S$, let $f(X)$ denote the dimensionality of the linear subspace spanned by the subspaces in X.

$S = \{ s_1, s_2, s_3 \}$

$s_1 = \{1, 2, 4\}$ $\quad s_2 = \{4, 5, 6\}$ $\quad s_3 = \{3, 5, 7\}$

$X \subseteq S$ $\quad X = \{ s_1, s_3 \}$

$f(X) = r(1, 2, 4, 3, 5, 7) = 6$

Polymatroid rank function

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

Polymatroid rank function

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

$$f(X) = r(\cup_{s \in X} X_s) \tag{4.10}$$

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

$$f(X) = r(\cup_{s \in X} X_s) \tag{4.10}$$

  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 $A \subseteq B$).

Polymatroid rank function

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

$$f(X) = r(\cup_{s \in X} X_s) \tag{4.10}$$

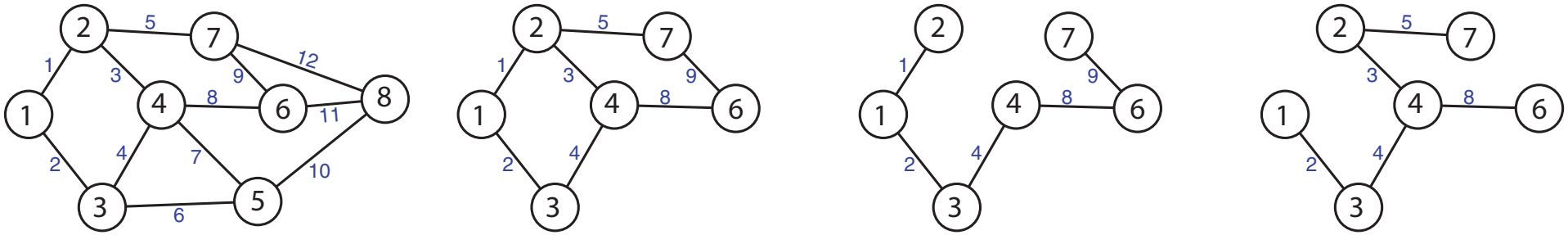
  we have that f is submodular, and is known to be a polymatroid rank function.
- In general (as we will see) polymatroid rank functions are submodular, normalized $f(\emptyset) = 0$, and monotone non-decreasing ($f(A) \leq f(B)$ whenever $A \subseteq B$).
- We use the term non-decreasing rather than increasing, the latter of which is strict (also so that a constant function isn't "increasing").

Spanning trees

- Let E be a set of edges of some graph $G = (V, E)$, and let $r(S)$ for $S \subseteq E$ be the maximum size (in terms of number of **edges**) spanning forest in the graph induced by edges S.

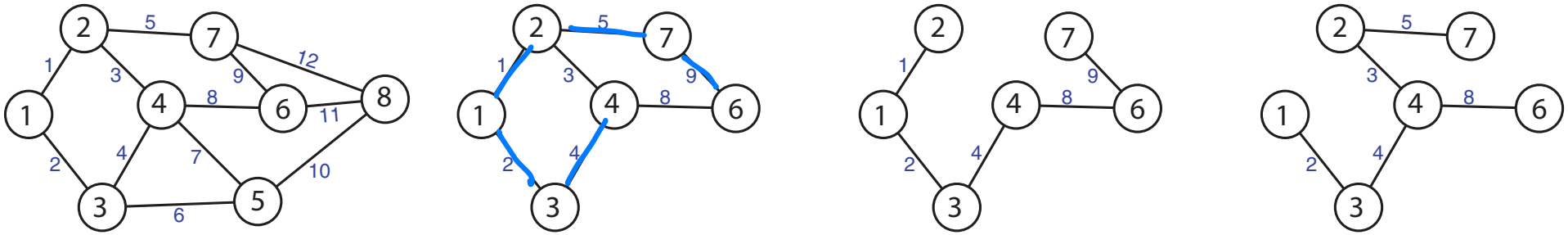
Spanning trees

- 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 graph induced by 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 graph induced by edges S.
- Example: Given $G = (V, E)$, $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$, $E = \{1, 2, \ldots, 12\}$. $S = \{1, 2, 3, 4, 5, 8, 9\} \subset E$. Two spanning trees have the same edge count (the rank of S).



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

Summing Submodular Functions

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

$$f : 2^E \to \mathbb{R} \text{ with } f(A) = f_1(A) + f_2(A) \tag{4.15}$$

is submodular.

Summing Submodular Functions

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

$$f : 2^E \to \mathbb{R} \text{ with } f(A) = f_1(A) + f_2(A) \tag{4.15}$$

is submodular. This follows easily since

$$\begin{aligned} f(A) + f(B) &= f_1(A) + f_2(A) + f_1(B) + f_2(B) && (4.16) \\ &\geq f_1(A \cup B) + f_2(A \cup B) + f_1(A \cap B) + f_2(A \cap B) && (4.17) \\ &= f(A \cup B) + f(A \cap B). && (4.18) \end{aligned}$$

I.e., it holds for each component of f in each term in the inequality.

Summing Submodular Functions

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

$$f : 2^E \to \mathbb{R} \text{ with } f(A) = f_1(A) + f_2(A) \tag{4.15}$$

is submodular.This follows easily since

$$\begin{aligned} f(A) + f(B) &= f_1(A) + f_2(A) + f_1(B) + f_2(B) && (4.16)\\ &\geq f_1(A \cup B) + f_2(A \cup B) + f_1(A \cap B) + f_2(A \cap B) && (4.17)\\ &= f(A \cup B) + f(A \cap B). && (4.18) \end{aligned}$$

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

$$\sum_i \alpha_i f_i(A) = f(A)$$

Summing Submodular and Modular Functions

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

Summing Submodular and Modular Functions

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

$$f : 2^E \to \mathbb{R} \text{ with } f(A) = f_1(A) - m(A) \tag{4.19}$$

is submodular (as is $f(A) = f_1(A) + m(A)$).

Summing Submodular and Modular Functions

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

$$f : 2^E \to \mathbb{R} \text{ with } f(A) = f_1(A) - m(A) \tag{4.19}$$

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

$$\begin{aligned} f(A) + f(B) &= f_1(A) - m(A) + f_1(B) - m(B) && (4.20)\\ &\geq f_1(A \cup B) - m(A \cup B) + f_1(A \cap B) - m(A \cap B) && (4.21)\\ &= f(A \cup B) + f(A \cap B). && (4.22) \end{aligned}$$

Summing Submodular and Modular Functions

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

$$f : 2^E \to \mathbb{R} \text{ with } f(A) = f_1(A) - m(A) \tag{4.19}$$

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

$$\begin{aligned} f(A) + f(B) &= f_1(A) - m(A) + f_1(B) - m(B) && (4.20)\\ &\geq f_1(A \cup B) - m(A \cup B) + f_1(A \cap B) - m(A \cap B) && (4.21)\\ &= f(A \cup B) + f(A \cap B). && (4.22) \end{aligned}$$

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

Restricting Submodular functions

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

$$f' : 2^E \to \mathbb{R} \text{ with } f'(A) \triangleq f(A \cap S) \tag{4.23}$$

is submodular.

Restricting Submodular functions

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

$$f' : 2^E \to \mathbb{R} \text{ with } f'(A) \triangleq f(A \cap S) \tag{4.23}$$

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

$$f' : 2^E \to \mathbb{R} \text{ with } f'(A) \triangleq f(A \cap S) \tag{4.23}$$

is submodular.

Proof.

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

$$f((A + v) \cap S) - f(A \cap S) \geq f((B + v) \cap S) - f(B \cap S) \tag{4.24}$$

□

Restricting Submodular functions

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

$$f' : 2^E \to \mathbb{R} \text{ with } f'(A) \triangleq f(A \cap S) \tag{4.23}$$

is submodular.

Proof.

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

$$f((A + v) \cap S) - f(A \cap S) \geq f((B + v) \cap S) - f(B \cap S) \tag{4.24}$$

If $v \notin S$, then both differences on each size are zero.

□

Restricting Submodular functions

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

$$f' : 2^E \to \mathbb{R} \text{ with } f'(A) \triangleq f(A \cap S) \tag{4.23}$$

is submodular.

Proof.

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

$$f((A + v) \cap S) - f(A \cap S) \geq f((B + v) \cap S) - f(B \cap S) \tag{4.24}$$

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

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

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 $S_1, S_2 \subseteq V$ be two arbitrary fixed sets. Then

$$f : 2^V \to \mathbb{R} \text{ with } f(A) = f_1(A \cap S_1) + f_2(A \cap S_2) \tag{4.26}$$

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

Summing Restricted Submodular Functions

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

$$f : 2^V \to \mathbb{R} \text{ with } f(A) = f_1(A \cap S_1) + f_2(A \cap S_2) \tag{4.26}$$

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 $f_C : \cancel{2^V \to \mathbb{R}}$ be a submodular function. Then

$f_C : 2^C \to \mathbb{R}$

$$f : 2^V \to \mathbb{R} \text{ with } f(A) = \sum_{C \in \mathcal{C}} f_C(A \cap C) \tag{4.27}$$

is submodular.

$$f(A) = \sum_{e=(u,v) \in E(G)} f_e(A \cap \{u,v\})$$

Summing Restricted Submodular Functions

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

$$f : 2^V \to \mathbb{R} \text{ with } f(A) = f_1(A \cap S_1) + f_2(A \cap S_2) \tag{4.26}$$

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

$$f : 2^V \to \mathbb{R} \text{ with } f(A) = \sum_{C \in \mathcal{C}} f_C(A \cap C) \tag{4.27}$$

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.

Max - normalized

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

$$f(A) = \max_{j \in A} c_j \tag{4.28}$$

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

Proof.

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

$$\max_{j \in A} c_j + \max_{j \in B} c_j \geq \max_{j \in A \cup B} c_j + \max_{j \in A \cap B} c_j \tag{4.29}$$

which follows since we have that

$$\max(\max_{j \in A} c_j, \max_{j \in B} c_j) = \max_{j \in A \cup B} c_j \tag{4.30}$$

and

$$\min(\max_{j \in A} c_j, \max_{j \in B} c_j) \geq \max_{j \in A \cap B} c_j \tag{4.31}$$

□

Max

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

$$f(A) = \max_{j \in A} c_j \tag{4.32}$$

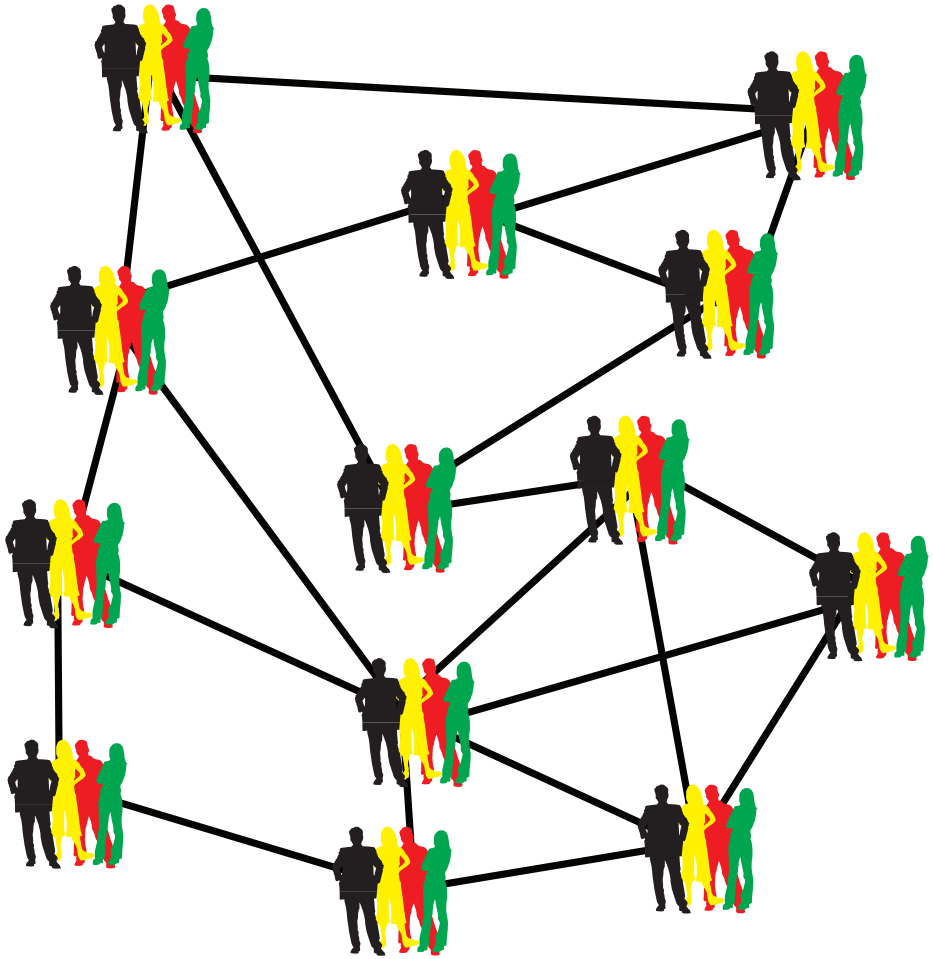
is submodular, where we take $f(\emptyset) \leq \min_j c_j$ (so the function need not be normalized).

Proof.

The proof is identical to the normalized case. □

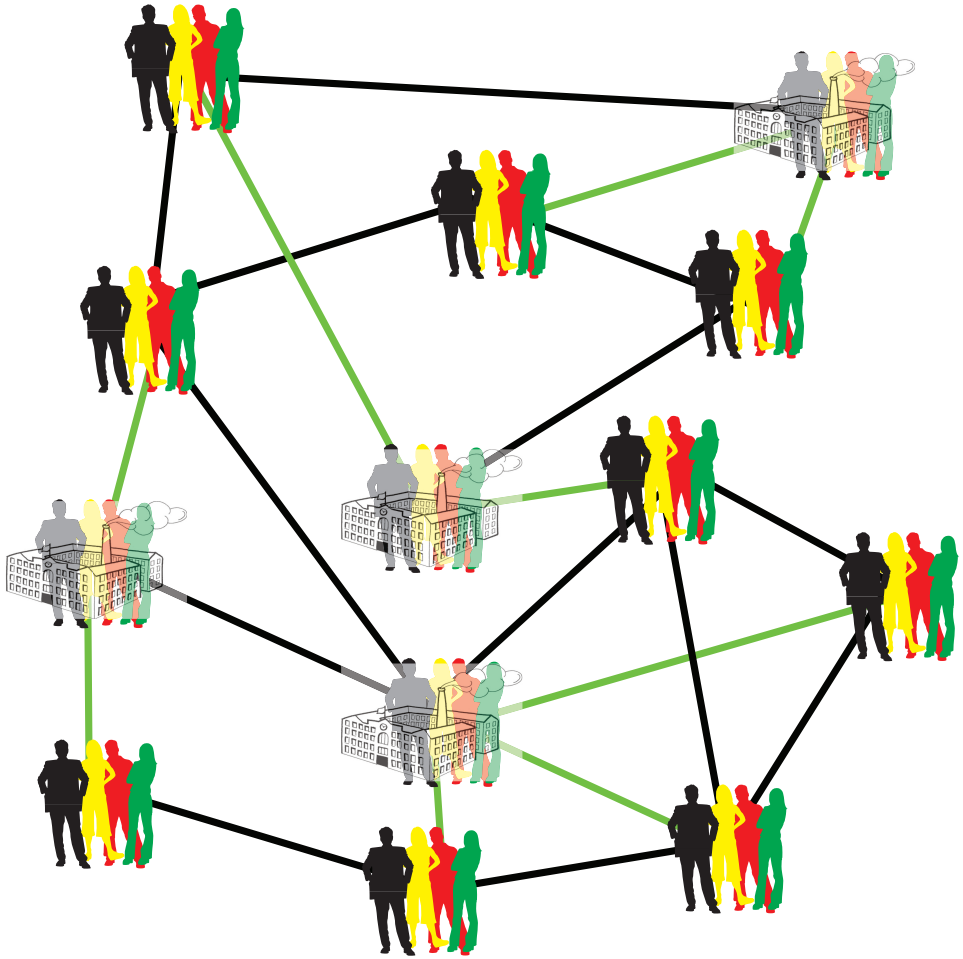
Facility/Plant Location (uncapacitated)

- 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.



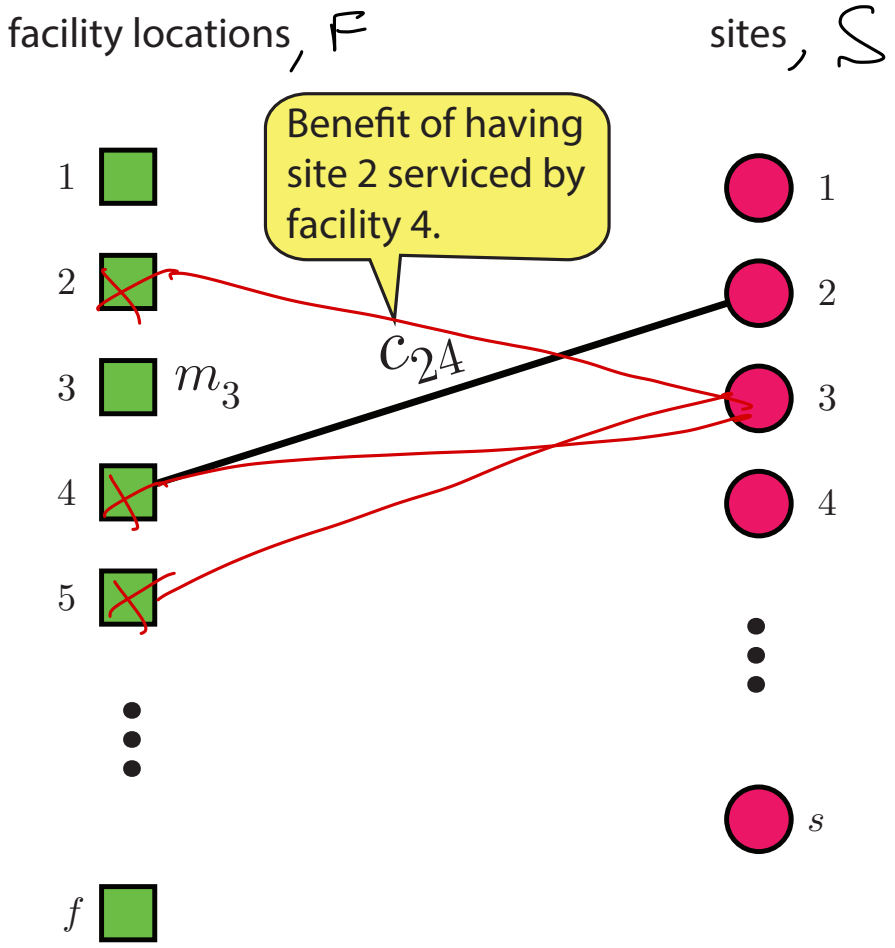
Facility/Plant Location (uncapacitated)

- 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 $c : E \to \mathbb{R}_+$ is the benefit of a given pair.
- Facility location function has form:

$A \subseteq F$,

$$f(A) = \sum_{i \in S} \max_{j \in A} c_{ij}. \qquad (4.33)$$



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.

Facility/Plant Location (uncapacitated) w. plant benefits

- Let $F = \{1, \dots, f\}$ be a set of possible factory/plant locations for facilities to be built.
- $S = \{1, \dots, s\}$ is a set of sites (e.g., cities, clients) needing service.

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" or "value" (e.g., $1/c_{ij}$ is the cost) of servicing site i with facility location j.

Facility/Plant Location (uncapacitated) w. plant benefits

- Let $F = \{1, \dots, f\}$ be a set of possible factory/plant locations for facilities to be built.
- $S = \{1, \dots, s\}$ is a set of sites (e.g., cities, clients) needing service.
- Let c_{ij} be the "benefit" or "value" (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" or "value" (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.

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

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

Facility/Plant Location (uncapacitated) w. plant benefits

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

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

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

Facility Location

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

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

is submodular.

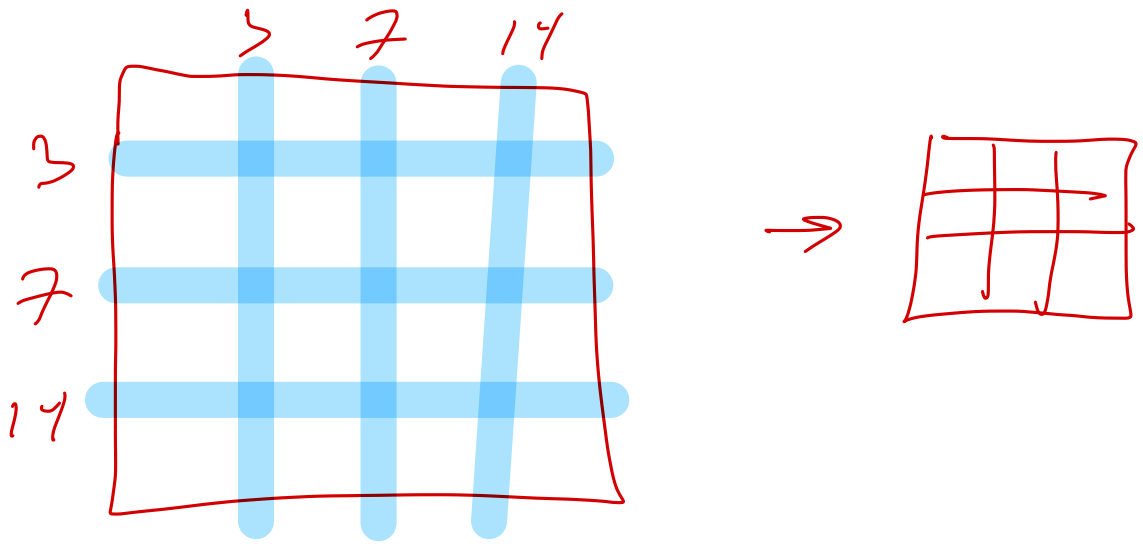
Proof.

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

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

Log Determinant

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



Log Determinant

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

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

Log Determinant

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

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

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

Log Determinant

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

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

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

Proof of submodularity of the logdet function.

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

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

...

Log Determinant

$I(X_A; X_B \mid X_C) \geq 0$

...cont.

$H(X_A, X_C) + H(X_B, X_C) - H(X_A, X_B, X_C) - H(X_C) \geq 0$

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

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

and in particular, for a variable subset A,

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

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

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

where $m(A)$ is a modular function. □

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

Summary: Properties so far

- Cover functions $f(A) = w(\bigcup_{a \in A} U_a)$ are submodular.

Summary: Properties so far

- Cover functions $f(A) = w(\bigcup_{a \in A} U_a)$ are submodular.
- SCCM is submodular $f(A) = \sum_{u \in U} \phi_u(m_u(A))$ where m_u is a non-negative modular and ϕ_u is concave.

Summary: Properties so far

- Cover functions $f(A) = w(\bigcup_{a \in A} U_a)$ are submodular.
- SCCM is submodular $f(A) = \sum_{u \in U} \phi_u(m_u(A))$ where m_u is a non-negative modular and ϕ_u is concave.
- max is submodular $f(A) = \max_{j \in A} c_j$, as is facility location $f(A) = \sum_{u \in U} \max_{a \in A} s_{a,u}$.

Summary: Properties so far

- Cover functions $f(A) = w(\bigcup_{a \in A} U_a)$ are submodular.
- SCCM is submodular $f(A) = \sum_{u \in U} \phi_u(m_u(A))$ where m_u is a non-negative modular and ϕ_u is concave.
- max is submodular $f(A) = \max_{j \in A} c_j$, as is facility location $f(A) = \sum_{u \in U} \max_{a \in A} s_{a,u}$.
- Matrix rank function is submodular.

Summary: Properties so far

- Cover functions $f(A) = w(\bigcup_{a \in A} U_a)$ are submodular.
- SCCM is submodular $f(A) = \sum_{u \in U} \phi_u(m_u(A))$ where m_u is a non-negative modular and ϕ_u is concave.
- max is submodular $f(A) = \max_{j \in A} c_j$, as is facility location $f(A) = \sum_{u \in U} \max_{a \in A} s_{a,u}$.
- Matrix rank function is submodular.
- Log determinant $f(A) = \log \det(\mathbf{\Sigma}_A)$ submodular for p.d. Σ.

Summary: Properties so far

- Cover functions $f(A) = w(\bigcup_{a \in A} U_a)$ are submodular.
- SCCM is submodular $f(A) = \sum_{u \in U} \phi_u(m_u(A))$ where m_u is a non-negative modular and ϕ_u is concave.
- max is submodular $f(A) = \max_{j \in A} c_j$, as is facility location $f(A) = \sum_{u \in U} \max_{a \in A} s_{a,u}$.
- Matrix rank function is submodular.
- Log determinant $f(A) = \log \det(\mathbf{\Sigma}_A)$ submodular for p.d. Σ.
- Matrix rank $r(A)$, dim. of space spanned by the vector set $\{x_a\}_{a \in A}$.

Summary: Properties so far

- Cover functions $f(A) = w(\bigcup_{a \in A} U_a)$ are submodular.
- SCCM is submodular $f(A) = \sum_{u \in U} \phi_u(m_u(A))$ where m_u is a non-negative modular and ϕ_u is concave.
- max is submodular $f(A) = \max_{j \in A} c_j$, as is facility location $f(A) = \sum_{u \in U} \max_{a \in A} s_{a,u}$.
- Matrix rank function is submodular.
- Log determinant $f(A) = \log \det(\mathbf{\Sigma}_A)$ submodular for p.d. Σ.
- Matrix rank $r(A)$, dim. of space spanned by the vector set $\{x_a\}_{a \in A}$.
- Graph cut, set cover, and incidence functions,

Summary: Properties so far

- Cover functions $f(A) = w(\bigcup_{a \in A} U_a)$ are submodular.
- SCCM is submodular $f(A) = \sum_{u \in U} \phi_u(m_u(A))$ where m_u is a non-negative modular and ϕ_u is concave.
- max is submodular $f(A) = \max_{j \in A} c_j$, as is facility location $f(A) = \sum_{u \in U} \max_{a \in A} s_{a,u}$.
- Matrix rank function is submodular.
- Log determinant $f(A) = \log \det(\mathbf{\Sigma}_A)$ submodular for p.d. Σ.
- Matrix rank $r(A)$, dim. of space spanned by the vector set $\{x_a\}_{a \in A}$.
- Graph cut, set cover, and incidence functions,
- quadratics with non-positive off-diagonals $f(X) = m^\intercal \mathbf{1}_X + \frac{1}{2} \mathbf{1}_X^\intercal \mathbf{M} \mathbf{1}_X$.

Summary: Properties so far

- Cover functions $f(A) = w(\bigcup_{a \in A} U_a)$ are submodular.
- SCCM is submodular $f(A) = \sum_{u \in U} \phi_u(m_u(A))$ where m_u is a non-negative modular and ϕ_u is concave.
- max is submodular $f(A) = \max_{j \in A} c_j$, as is facility location $f(A) = \sum_{u \in U} \max_{a \in A} s_{a,u}$.
- Matrix rank function is submodular.
- Log determinant $f(A) = \log \det(\mathbf{\Sigma}_A)$ submodular for p.d. Σ.
- Matrix rank $r(A)$, dim. of space spanned by the vector set $\{x_a\}_{a \in A}$.
- Graph cut, set cover, and incidence functions,
- quadratics with non-positive off-diagonals $f(X) = m^\intercal \mathbf{1}_X + \frac{1}{2} \mathbf{1}_X^\intercal \mathbf{M} \mathbf{1}_X$.
- Number connected components in induced graph $c(A)$, and interior edge function $E(S)$, is supermodular.

Summary: Properties so far

- Cover functions $f(A) = w(\bigcup_{a \in A} U_a)$ are submodular.
- SCCM is submodular $f(A) = \sum_{u \in U} \phi_u(m_u(A))$ where m_u is a non-negative modular and ϕ_u is concave.
- max is submodular $f(A) = \max_{j \in A} c_j$, as is facility location $f(A) = \sum_{u \in U} \max_{a \in A} s_{a,u}$.
- Matrix rank function is submodular.
- Log determinant $f(A) = \log \det(\mathbf{\Sigma}_A)$ submodular for p.d. Σ.
- Matrix rank $r(A)$, dim. of space spanned by the vector set $\{x_a\}_{a \in A}$.
- Graph cut, set cover, and incidence functions,
- quadratics with non-positive off-diagonals $f(X) = m^\intercal \mathbf{1}_X + \frac{1}{2}\mathbf{1}_X^\intercal \mathbf{M} \mathbf{1}_X$.
- Number connected components in induced graph $c(A)$, and interior edge function $E(S)$, is supermodular.
- Submodular plus modular is submodular, $f(A) = f'(A) + m(A)$.

Summary: Properties so far

- Cover functions $f(A) = w(\bigcup_{a \in A} U_a)$ are submodular.
- SCCM is submodular $f(A) = \sum_{u \in U} \phi_u(m_u(A))$ where m_u is a non-negative modular and ϕ_u is concave.
- max is submodular $f(A) = \max_{j \in A} c_j$, as is facility location $f(A) = \sum_{u \in U} \max_{a \in A} s_{a,u}$.
- Matrix rank function is submodular.
- Log determinant $f(A) = \log \det(\mathbf{\Sigma}_A)$ submodular for p.d. Σ.
- Matrix rank $r(A)$, dim. of space spanned by the vector set $\{x_a\}_{a \in A}$.
- Graph cut, set cover, and incidence functions,
- quadratics with non-positive off-diagonals $f(X) = m^\intercal \mathbf{1}_X + \frac{1}{2}\mathbf{1}_X^\intercal \mathbf{M} \mathbf{1}_X$.
- Number connected components in induced graph $c(A)$, and interior edge function $E(S)$, is supermodular.
- Submodular plus modular is submodular, $f(A) = f'(A) + m(A)$.
- Complementation: $f'(A) = f(V \setminus A)$ is submodular if f is submodular (same for supermodular, modular).

Summary: Properties so far

- Cover functions $f(A) = w(\bigcup_{a \in A} U_a)$ are submodular.
- SCCM is submodular $f(A) = \sum_{u \in U} \phi_u(m_u(A))$ where m_u is a non-negative modular and ϕ_u is concave.
- max is submodular $f(A) = \max_{j \in A} c_j$, as is facility location $f(A) = \sum_{u \in U} \max_{a \in A} s_{a,u}$.
- Matrix rank function is submodular.
- Log determinant $f(A) = \log\det(\mathbf{\Sigma}_A)$ submodular for p.d. Σ.
- Matrix rank $r(A)$, dim. of space spanned by the vector set $\{x_a\}_{a \in A}$.
- Graph cut, set cover, and incidence functions,
- quadratics with non-positive off-diagonals $f(X) = m^\intercal \mathbf{1}_X + \frac{1}{2}\mathbf{1}_X^\intercal \mathbf{M} \mathbf{1}_X$.
- Number connected components in induced graph $c(A)$, and interior edge function $E(S)$, is supermodular.
- Submodular plus modular is submodular, $f(A) = f'(A) + m(A)$.
- Complementation: $f'(A) = f(V \setminus A)$ is submodular if f is submodular (same for supermodular, modular).
- Conix mixture: $\alpha_i \geq 0$, $f_i : 2^V \to \mathbb{R}$ submodular, then so is $\sum_i \alpha_i f_i$.

Summary: Properties so far

- Cover functions $f(A) = w(\bigcup_{a \in A} U_a)$ are submodular.
- SCCM is submodular $f(A) = \sum_{u \in U} \phi_u(m_u(A))$ where m_u is a non-negative modular and ϕ_u is concave.
- max is submodular $f(A) = \max_{j \in A} c_j$, as is facility location $f(A) = \sum_{u \in U} \max_{a \in A} s_{a,u}$.
- Matrix rank function is submodular.
- Log determinant $f(A) = \log \det(\mathbf{\Sigma}_A)$ submodular for p.d. Σ.
- Matrix rank $r(A)$, dim. of space spanned by the vector set $\{x_a\}_{a \in A}$.
- Graph cut, set cover, and incidence functions,
- quadratics with non-positive off-diagonals $f(X) = m^\intercal \mathbf{1}_X + \frac{1}{2}\mathbf{1}_X^\intercal \mathbf{M} \mathbf{1}_X$.
- Number connected components in induced graph $c(A)$, and interior edge function $E(S)$, is supermodular.
- Submodular plus modular is submodular, $f(A) = f'(A) + m(A)$.
- Complementation: $f'(A) = f(V \setminus A)$ is submodular if f is submodular (same for supermodular, modular).
- Conix mixture: $\alpha_i \geq 0$, $f_i : 2^V \to \mathbb{R}$ submodular, then so is $\sum_i \alpha_i f_i$.
- Restrictions preserve submodularity: $f'(A) = f(A \cap S)$

Concave over non-negative modular

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

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

then f is submodular.

Concave over non-negative modular

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

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

then f is submodular.

Proof.

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

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

□

A form of converse is true as well.

Concave composed with non-negative modular

Theorem 4.5.1

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

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

- If ϕ is non-decreasing concave w. $\phi(0) = 0$, then f is polymatroidal.

Concave composed with non-negative modular

Theorem 4.5.1

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

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

- If ϕ is non-decreasing concave w. $\phi(0) = 0$, then f is polymatroidal.
- Sums of concave over modular functions are submodular

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

Concave composed with non-negative modular

Theorem 4.5.1

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

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

- If ϕ is non-decreasing concave w. $\phi(0) = 0$, then f is polymatroidal.
- Sums of concave over modular functions are submodular

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

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

Concave composed with non-negative modular

Theorem 4.5.1

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

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

- If ϕ is non-decreasing concave w. $\phi(0) = 0$, then f is polymatroidal.
- Sums of concave over modular functions are submodular

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

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

Monotonicity

Definition 4.5.2

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

Monotonicity

Definition 4.5.2

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

Definition 4.5.3

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

Composition of non-decreasing submodular and non-decreasing concave

Theorem 4.5.4

Given two functions, one defined on sets

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

and another continuous valued one:

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

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

Monotone difference of two functions

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

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

is submodular.

Proof.

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

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

or

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

the result (Equation 4.46 being submodular) follows since

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

...

Monotone difference of two functions

...cont.

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

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

Monotone difference of two functions

...cont.

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

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

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

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

□

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

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

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

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

is submodular.

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

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

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

is submodular.

Proof.

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

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

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

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

is submodular.

Proof.

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

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

More on Min - the saturate trick

- minimax facility location is similar to the following maximin function (a form of "robust facility location"): $h(A) = \min_{v \in V} \max_{a \in A} s(i, a)$ and the goal is to maximize this $\max_{A \subseteq V : |A| \leq k} h(A)$. h therefore is the min of a set of submodular functions.

More on Min - the saturate trick

- minimax facility location is similar to the following maximin function (a form of "robust facility location"): $h(A) = \min_{v \in V} \max_{a \in A} s(i, a)$ and the goal is to maximize this $\max_{A \subseteq V : |A| \leq k} h(A)$. h therefore is the min of a set of submodular functions.
- In general, the minimum of two submodular functions is not submodular (unlike concave functions, closed under min).

More on Min - the saturate trick

- minimax facility location is similar to the following maximin function (a form of "robust facility location"): $h(A) = \min_{v \in V} \max_{a \in A} s(i, a)$ and the goal is to maximize this $\max_{A \subseteq V : |A| \leq k} h(A)$. h therefore is the min of a set of submodular functions.
- In general, the minimum of two submodular functions is not submodular (unlike concave functions, closed under min).
- However, when wishing to maximize two monotone non-decreasing submodular functions f, g, we can define function $h_\alpha : 2^V \to \mathbb{R}$ as

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

then h_α is submodular, and $h_\alpha(A) \geq \alpha$ if and only if both $f(A) \geq \alpha$ and $g(A) \geq \alpha$, for constant $\alpha \in \mathbb{R}$.

More on Min - the saturate trick

- minimax facility location is similar to the following maximin function (a form of "robust facility location"): $h(A) = \min_{v \in V} \max_{a \in A} s(i, a)$ and the goal is to maximize this $\max_{A \subseteq V : |A| \leq k} h(A)$. h therefore is the min of a set of submodular functions.
- In general, the minimum of two submodular functions is not submodular (unlike concave functions, closed under min).
- However, when wishing to maximize two monotone non-decreasing submodular functions f, g, we can define function $h_\alpha : 2^V \to \mathbb{R}$ as

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

  then h_α is submodular, and $h_\alpha(A) \geq \alpha$ if and only if both $f(A) \geq \alpha$ and $g(A) \geq \alpha$, for constant $\alpha \in \mathbb{R}$.
- Useful in applications. Like DS functions, another instance of a submodular surrogate (where we take a non-submodular problem and find a submodular one that can tell us something about it).

Arbitrary functions: difference between submodular funcs.

Theorem 4.5.5

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

Proof.

Let h be given and arbitrary, and define:

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

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

Arbitrary functions: difference between submodular funcs.

Theorem 4.5.5

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

Proof.

Let h be given and arbitrary, and define:

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

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

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

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

Arbitrary functions as difference between submodular funcs.

...cont.

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

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

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

□

Gain

- We often wish to express the gain of an item $j \in V$ in context A, namely $f(A \cup \{j\}) - f(A)$.

Gain

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

$$f(A \cup \{j\}) - f(A) \triangleq \rho_j(A) \tag{4.57}$$

$$\triangleq \rho_A(j) \tag{4.58}$$

$$\triangleq \nabla_j f(A) \tag{4.59}$$

$$\triangleq f(\{j\}|A) \tag{4.60}$$

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

Gain

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

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

- We'll use $f(j|A)$.

Gain

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

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

- We'll use $f(j|A)$.
- diminishing returns 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 $A \subseteq B$ (i.e., further conditioning reduces valuation).

Gain Notation

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

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

So when j is any singleton

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

Gain Notation

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

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

So when j is any singleton

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

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

Totally normalized functions

- Any normalized submodular function g (even non-monotone) can be represented as a sum of a polymatroid (normalized monotone non-decreasing submodular) function $\bar{g}$ and a modular function m_g.

Totally normalized functions

- Any normalized submodular function g (even non-monotone) can be represented as a sum of a polymatroid (normalized monotone non-decreasing submodular) function $\bar{g}$ and a modular function m_g.
- E.g., $g(A) = [g(A) + \alpha|A|] - \alpha|A|$, $\alpha \geq |\min_{v, A \subseteq V \setminus v} f(v|A)|$.

Totally normalized functions

- Any normalized submodular function g (even non-monotone) can be represented as a sum of a polymatroid (normalized monotone non-decreasing submodular) function $\bar{g}$ and a modular function m_g.
- E.g., $g(A) = [g(A) + \alpha|A|] - \alpha|A|$, $\alpha \geq |\min_{v, A \subseteq V \setminus v} f(v|A)|$.
- Given arbitrary normalized submodular $g : 2^V \to \mathbb{R}$, construct a function $\bar{g} : 2^V \to \mathbb{R}$ as follows:

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

where $m_g(A) \triangleq \sum_{a \in A} g(a|V \setminus \{a\})$ is a modular function.

Totally normalized functions

- Any normalized submodular function g (even non-monotone) can be represented as a sum of a polymatroid (normalized monotone non-decreasing submodular) function $\bar{g}$ and a modular function m_g.
- E.g., $g(A) = [g(A) + \alpha|A|] - \alpha|A|$, $\alpha \geq |\min_{v, A \subseteq V \setminus v} f(v|A)|$.
- Given arbitrary normalized submodular $g : 2^V \to \mathbb{R}$, construct a function $\bar{g} : 2^V \to \mathbb{R}$ as follows:

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

  where $m_g(A) \triangleq \sum_{a \in A} g(a|V \setminus \{a\})$ is a modular function.
- $\bar{g}$ is normalized since $\bar{g}(\emptyset) = 0$.

Totally normalized functions

- Any normalized submodular function g (even non-monotone) can be represented as a sum of a polymatroid (normalized monotone non-decreasing submodular) function $\bar{g}$ and a modular function m_g.
- E.g., $g(A) = [g(A) + \alpha|A|] - \alpha|A|$, $\alpha \geq |\min_{v, A \subseteq V \setminus v} f(v|A)|$.
- Given arbitrary normalized submodular $g : 2^V \to \mathbb{R}$, construct a function $\bar{g} : 2^V \to \mathbb{R}$ as follows:

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

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

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

Totally normalized functions

- Any normalized submodular function g (even non-monotone) can be represented as a sum of a polymatroid (normalized monotone non-decreasing submodular) function $\bar{g}$ and a modular function m_g.
- E.g., $g(A) = [g(A) + \alpha|A|] - \alpha|A|$, $\alpha \geq |\min_{v, A \subseteq V \setminus v} f(v|A)|$.
- Given arbitrary normalized submodular $g : 2^V \to \mathbb{R}$, construct a function $\bar{g} : 2^V \to \mathbb{R}$ as follows:

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

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

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

- $\bar{g}$ is called the totally normalized version of g.

Totally normalized functions

- Any normalized submodular function g (even non-monotone) can be represented as a sum of a polymatroid (normalized monotone non-decreasing submodular) function $\bar{g}$ and a modular function m_g.
- E.g., $g(A) = [g(A) + \alpha|A|] - \alpha|A|$, $\alpha \geq |\min_{v, A \subseteq V \setminus v} f(v|A)|$.
- Given arbitrary normalized submodular $g : 2^V \to \mathbb{R}$, construct a function $\bar{g} : 2^V \to \mathbb{R}$ as follows:

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

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

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

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

Arbitrary function as difference between two polymatroids

- Any normalized function h (i.e., $h(\emptyset) = 0$) can be represented as a difference not only between submodular, but between polymatroid (normalized monotone non-decreasing submodular) functions.

Arbitrary function as difference between two polymatroids

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

Arbitrary function as difference between two polymatroids

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

$$h = f - g = \bar{f} + m_f - (\bar{g} + m_g) \tag{4.66}$$
$$= \bar{f} - \bar{g} + (m_f - m_g) \tag{4.67}$$
$$= \bar{f} - \bar{g} + m_{f-h} \tag{4.68}$$
$$= \bar{f} + m^+_{f-g} - (\bar{g} + (-m_{f-g})^+) \tag{4.69}$$

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

Arbitrary function as difference between two polymatroids

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

$$h = f - g = \bar{f} + m_f - (\bar{g} + m_g) \tag{4.66}$$
$$= \bar{f} - \bar{g} + (m_f - m_g) \tag{4.67}$$
$$= \bar{f} - \bar{g} + m_{f-h} \tag{4.68}$$
$$= \bar{f} + m^+_{f-g} - (\bar{g} + (-m_{f-g})^+) \tag{4.69}$$

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

- Both $\bar{f} + m^+_{f-g}$ and $\bar{g} + (-m_{f-g})^+$ are polymatroid functions!

Arbitrary function as difference between two polymatroids

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

$$h = f - g = \bar{f} + m_f - (\bar{g} + m_g) \tag{4.66}$$
$$= \bar{f} - \bar{g} + (m_f - m_g) \tag{4.67}$$
$$= \bar{f} - \bar{g} + m_{f-h} \tag{4.68}$$
$$= \bar{f} + m^+_{f-g} - (\bar{g} + (-m_{f-g})^+) \tag{4.69}$$

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

- Both $\bar{f} + m^+_{f-g}$ and $\bar{g} + (-m_{f-g})^+$ are polymatroid functions!
- Thus, any function can be expressed as a difference between two, not only submodular (DS), but polymatroid functions.

Two Equivalent Submodular Definitions

Definition 4.6.1 (submodular concave)

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

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B) \tag{4.7}$$

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

Definition 4.6.2 (diminishing returns)

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

$$f(A \cup \{v\}) - f(A) \geq f(B \cup \{v\}) - f(B) \tag{4.8}$$

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

• Gain notation: Define $f(v|A) \triangleq f(A + v) - f(A)$. Then function f is submodular if $f(v|A) \geq f(v|B)$ for all $A \subseteq B \subseteq V \setminus \{v\}$, $v \in V$.

Submodular Definition: Group Diminishing Returns

An alternate and equivalent definition is:

Definition 4.6.1 (group diminishing returns)

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

$$f(A \cup C) - f(A) \geq f(B \cup C) - f(B) \tag{4.70}$$

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

Submodular Definition Basic Equivalencies

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

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 $f(S) + f(T) \geq f(S \cup T) + f(S \cap T)$.

$\square$

Submodular Concave $\Rightarrow$ Diminishing Returns

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

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

$$f(A + v) + f(B) \geq f(B + v) + f(A) \tag{4.71}$$

$\square$

Submodular Concave $\Rightarrow$ Diminishing Returns

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

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

$$f(A + v) + f(B) \geq f(B + v) + f(A) \tag{4.71}$$

- Rearranging, we have

$$f(A + v) - f(A) \geq f(B + v) - f(B) \tag{4.72}$$

$\square$

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

$$f(A \cup C) - f(A) \tag{4.73}$$

$$= f(A \cup C) - \sum_{i=1}^{k-1} \Big(f(A \cup \{c_1, \dots, c_i\}) - f(A \cup \{c_1, \dots, c_i\}) \Big) - f(A) \tag{4.74}$$

$$= \sum_{i=1}^{k} \Big(f(A \cup \{c_1 \dots c_i\}) - f(A \cup \{c_1 \dots c_{i-1}\}) \Big) = \sum_{i=1}^{k} f(c_i | A \cup \{c_1 \dots c_{i-1}\}) \tag{4.75}$$

$$\geq \sum_{i=1}^{k} f(c_i | B \cup \{c_1 \dots c_{i-1}\}) = \sum_{i=1}^{k} \Big(f(B \cup \{c_1 \dots c_i\}) - f(B \cup \{c_1 \dots c_{i-1}\}) \Big) \tag{4.76}$$

$$= f(B \cup C) - \sum_{i=1}^{k-1} \Big(f(B \cup \{c_1, \dots, c_i\}) - f(B \cup \{c_1, \dots, c_i\}) \Big) - f(B) \tag{4.77}$$

$$= f(B \cup C) - f(B) \tag{4.78}$$

$\square$

Group Diminishing Returns $\Rightarrow$ Submodular Concave

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

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

$$f(A' + C) - f(A') \geq f(B' + C) - f(B') \tag{4.79}$$

giving

$$f(A' + C) + f(B') \geq f(B' + C) + f(A') \tag{4.80}$$

or

$$f(A \cap B + A \setminus B) + f(B) \geq f(B + A \setminus B) + f(A \cap B) \tag{4.81}$$

which is the same as the submodular concave condition

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B) \tag{4.82}$$

Submodular Definition: Four Points

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

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

$$f(A \cup \{a\}) + f(A \cup \{b\}) \geq f(A \cup \{a, b\}) + f(A) \tag{4.83}$$

Submodular Definition: Four Points

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

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

$$f(A \cup \{a\}) + f(A \cup \{b\}) \geq f(A \cup \{a, b\}) + f(A) \tag{4.83}$$

This follows immediately from diminishing returns.

Submodular Definition: Four Points

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

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

$$f(A \cup \{a\}) + f(A \cup \{b\}) \geq f(A \cup \{a, b\}) + f(A) \tag{4.83}$$

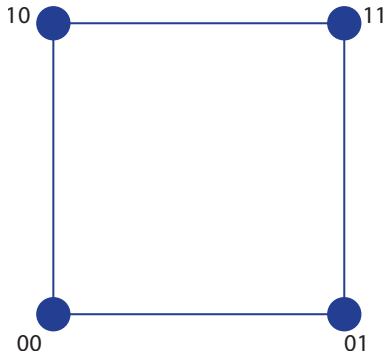
This follows immediately from diminishing returns. To achieve diminishing returns, assume $A \subset B$ with $B \setminus A = \{b_1, b_2, \dots, b_k\}$. Then

$$\begin{aligned}
f(A + a) - f(A) &\geq f(A + b_1 + a) - f(A + b_1) && (4.84)\\
&\geq f(A + b_1 + b_2 + a) - f(A + b_1 + b_2) && (4.85)\\
&\geq \dots && (4.86)\\
&\geq f(A + b_1 + \dots + b_k + a) - f(A + b_1 + \dots + b_k) && (4.87)\\
&= f(B + a) - f(B) && (4.88)
\end{aligned}$$

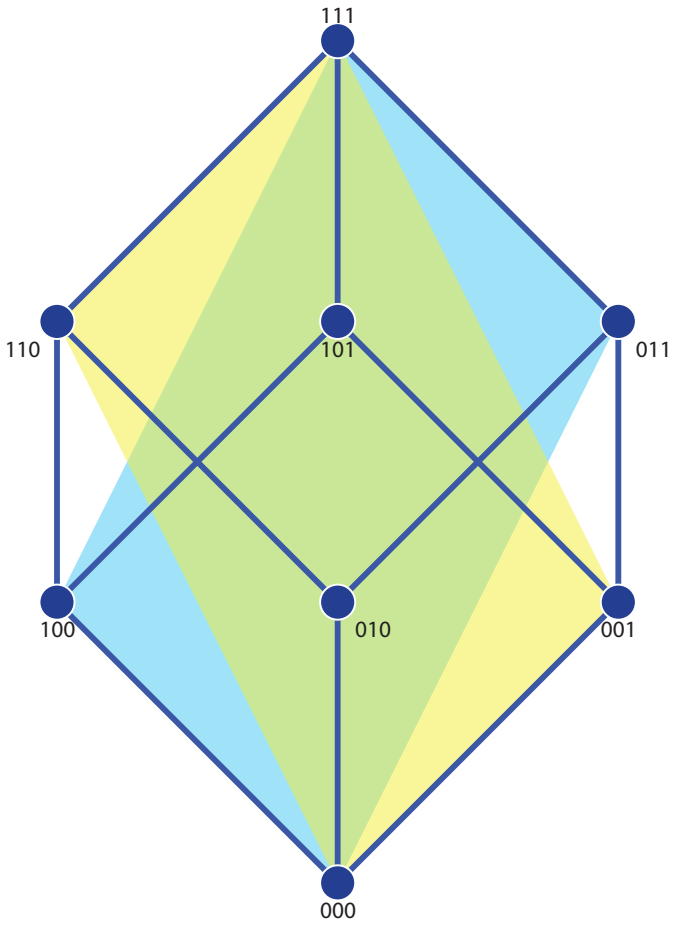
The Submodular Square, and Hypercube Vertices

We can test submodularity via values on vertices of hypercube.

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



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



How many inequalities of form $f(A) + f(B) \geq f(A \cup B) + f(A \cap B)$?

Submodular Concave $\equiv$ Diminishing Returns, in one slide.

Theorem 4.6.3

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

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

if and only if

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

Proof.

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

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

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

Applying telescoping summation to both sides, we get:

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

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

$\square$

Many (Equivalent) Definitions of Submodularity

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B), \quad \forall A, B \subseteq V \tag{4.94}$$

Many (Equivalent) Definitions of Submodularity

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B), \;\; \forall A, B \subseteq V \tag{4.94}$$

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

Many (Equivalent) Definitions of Submodularity

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B), \quad \forall A, B \subseteq V \tag{4.94}$$

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

$$f(C|S) \geq f(C|T), \forall S \subseteq T \subseteq V, \text{ with } C \subseteq V \setminus T \tag{4.96}$$

Many (Equivalent) Definitions of Submodularity

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B), \quad \forall A, B \subseteq V \tag{4.94}$$

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

$$f(C|S) \geq f(C|T), \forall S \subseteq T \subseteq V, \text{ with } C \subseteq V \setminus T \tag{4.96}$$

$$f(j|S) \geq f(j|S \cup \{k\}), \ \forall S \subseteq V \text{ with } j \in V \setminus (S \cup \{k\}) \tag{4.97}$$

Many (Equivalent) Definitions of Submodularity

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B), \ \ \forall A, B \subseteq V \tag{4.94}$$

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

$$f(C|S) \geq f(C|T), \forall S \subseteq T \subseteq V, \text{ with } C \subseteq V \setminus T \tag{4.96}$$

$$f(j|S) \geq f(j|S \cup \{k\}), \ \forall S \subseteq V \text{ with } j \in V \setminus (S \cup \{k\}) \tag{4.97}$$

$$f(A \cup B|A \cap B) \leq f(A|A \cap B) + f(B|A \cap B), \ \ \forall A, B \subseteq V \tag{4.98}$$

Many (Equivalent) Definitions of Submodularity

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B), \ \ \forall A, B \subseteq V \tag{4.94}$$

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

$$f(C|S) \geq f(C|T), \forall S \subseteq T \subseteq V, \text{ with } C \subseteq V \setminus T \tag{4.96}$$

$$f(j|S) \geq f(j|S \cup \{k\}), \ \forall S \subseteq V \text{ with } j \in V \setminus (S \cup \{k\}) \tag{4.97}$$

$$f(A \cup B|A \cap B) \leq f(A|A \cap B) + f(B|A \cap B), \ \ \forall A, B \subseteq V \tag{4.98}$$

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

Many (Equivalent) Definitions of Submodularity

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B), \ \ \forall A, B \subseteq V \tag{4.94}$$

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

$$f(C|S) \geq f(C|T), \forall S \subseteq T \subseteq V, \text{ with } C \subseteq V \setminus T \tag{4.96}$$

$$f(j|S) \geq f(j|S \cup \{k\}), \ \forall S \subseteq V \text{ with } j \in V \setminus (S \cup \{k\}) \tag{4.97}$$

$$f(A \cup B|A \cap B) \leq f(A|A \cap B) + f(B|A \cap B), \ \ \forall A, B \subseteq V \tag{4.98}$$

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

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

Many (Equivalent) Definitions of Submodularity

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B), \quad \forall A, B \subseteq V \tag{4.94}$$

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

$$f(C|S) \geq f(C|T), \forall S \subseteq T \subseteq V, \text{ with } C \subseteq V \setminus T \tag{4.96}$$

$$f(j|S) \geq f(j|S \cup \{k\}), \ \forall S \subseteq V \text{ with } j \in V \setminus (S \cup \{k\}) \tag{4.97}$$

$$f(A \cup B|A \cap B) \leq f(A|A \cap B) + f(B|A \cap B), \quad \forall A, B \subseteq V \tag{4.98}$$

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

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

$$f(T) \leq f(S) - \sum_{j \in S \setminus T} f(j|S \setminus \{j\}) + \sum_{j \in T \setminus S} f(j|S \cap T) \ \forall S, T \subseteq V \tag{4.101}$$

Many (Equivalent) Definitions of Submodularity

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B), \quad \forall A, B \subseteq V \tag{4.94}$$

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

$$f(C|S) \geq f(C|T), \forall S \subseteq T \subseteq V, \text{ with } C \subseteq V \setminus T \tag{4.96}$$

$$f(j|S) \geq f(j|S \cup \{k\}), \ \forall S \subseteq V \text{ with } j \in V \setminus (S \cup \{k\}) \tag{4.97}$$

$$f(A \cup B|A \cap B) \leq f(A|A \cap B) + f(B|A \cap B), \quad \forall A, B \subseteq V \tag{4.98}$$

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

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

$$f(T) \leq f(S) - \sum_{j \in S \setminus T} f(j|S \setminus \{j\}) + \sum_{j \in T \setminus S} f(j|S \cap T) \ \forall S, T \subseteq V \tag{4.101}$$

$$f(T) \leq f(S) - \sum_{j \in S \setminus T} f(j|S \setminus \{j\}), \ \forall T \subseteq S \subseteq V \tag{4.102}$$

Equivalent Definitions of Submodularity

We've already seen that Eq. 4.94 $\equiv$ Eq. 4.95 $\equiv$ Eq. 4.96 $\equiv$ Eq. 4.97 $\equiv$ Eq. 4.98.

Equivalent Definitions of Submodularity

We've already seen that Eq. 4.94 $\equiv$ Eq. 4.95 $\equiv$ Eq. 4.96 $\equiv$ Eq. 4.97 $\equiv$ Eq. 4.98.

We next show that Eq. 4.97 $\Rightarrow$ Eq. 4.99 $\Rightarrow$ Eq. 4.100 $\Rightarrow$ Eq. 4.97.

Approach

To show these next results, we essentially first use:

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

and

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

Approach

To show these next results, we essentially first use:

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

and

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

leading to

$$f(T) + \text{lower-bound} \leq f(S) + \text{upper-bound} \tag{4.105}$$

or

$$f(T) \leq f(S) + \text{upper-bound} - \text{lower-bound} \tag{4.106}$$

Eq. 4.97 $\Rightarrow$ Eq. 4.99

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

$$f(S \cup T) - f(S) = \sum_{t=1}^{r} \Big(f(S \cup \{j_1, \dots, j_t\}) - f(S \cup \{j_1, \dots, j_{t-1}\}) \Big) \tag{4.107}$$

$$= \sum_{t=1}^{r} f(j_t | S \cup \{j_1, \dots, j_{t-1}\}) \leq \sum_{t=1}^{r} f(j_t | S) \tag{4.108}$$

$$= \sum_{j \in T \setminus S} f(j|S) \tag{4.109}$$

or

$$f(T|S) \leq \sum_{j \in T \setminus S} f(j|S) \tag{4.110}$$

Eq. 4.97 $\Rightarrow$ Eq. 4.99

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

$$f(S \cup T) - f(T) = \sum_{t=1}^{q} [f(T \cup \{k_1, \ldots, k_t\}) - f(T \cup \{k_1, \ldots, k_{t-1}\})] \tag{4.111}$$

$$= \sum_{t=1}^{q} f(k_t | T \cup \{k_1, \ldots, k_t\} \setminus \{k_t\}) \geq \sum_{t=1}^{q} f(k_t | T \cup S \setminus \{k_t\}) \tag{4.112}$$

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

Eq. 4.97 $\Rightarrow$ Eq. 4.99

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

$$f(T|S) = f(S \cup T) - f(S) \leq \sum_{j \in T \setminus S} f(j|S) \tag{4.114}$$

and the lower bound

$$f(S|T) = f(S \cup T) - f(T) \geq \sum_{j \in S \setminus T} f(j|S \cup T \setminus \{j\}) \tag{4.115}$$

This gives upper and lower bounds of the form

$$f(T) + \text{lower bound} \leq f(S \cup T) \leq f(S) + \text{upper bound}, \tag{4.116}$$

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

Eq. 4.99 $\Rightarrow$ Eq. 4.100

This follows immediately since if $S \subseteq T$, then $S \setminus T = \emptyset$, and the last term of Eq. 4.99 vanishes.

Many (Equivalent) Definitions of Submodularity

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B), \quad \forall A, B \subseteq V \tag{4.94}$$

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

$$f(C|S) \geq f(C|T), \forall S \subseteq T \subseteq V, \text{ with } C \subseteq V \setminus T \tag{4.96}$$

$$f(j|S) \geq f(j|S \cup \{k\}), \ \forall S \subseteq V \text{ with } j \in V \setminus (S \cup \{k\}) \tag{4.97}$$

$$f(A \cup B|A \cap B) \leq f(A|A \cap B) + f(B|A \cap B), \quad \forall A, B \subseteq V \tag{4.98}$$

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

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

$$f(T) \leq f(S) - \sum_{j \in S \setminus T} f(j|S \setminus \{j\}) + \sum_{j \in T \setminus S} f(j|S \cap T) \ \forall S, T \subseteq V \tag{4.101}$$

$$f(T) \leq f(S) - \sum_{j \in S \setminus T} f(j|S \setminus \{j\}), \ \forall T \subseteq S \subseteq V \tag{4.102}$$

Eq. 4.100 $\Rightarrow$ Eq. 4.97

Here, we set $T = S \cup \{j, k\}$, $j \notin S \cup \{k\}$ into Eq. 4.100 to obtain

$$
\begin{aligned}
f(S \cup \{j,k\}) &\leq f(S) + f(j|S) + f(k|S) && (4.117)\\
&= f(S) + f(S + \{j\}) - f(S) + f(S + \{k\}) - f(S) && (4.118)\\
&= f(S + \{j\}) + f(S + \{k\}) - f(S) && (4.119)\\
&= f(j|S) + f(S + \{k\}) && (4.120)
\end{aligned}
$$

giving

$$
\begin{aligned}
f(j|S \cup \{k\}) &= f(S \cup \{j,k\}) - f(S \cup \{k\}) && (4.121)\\
&\leq f(j|S) && (4.122)
\end{aligned}
$$

Submodular Concave

- Why do we call the $f(A) + f(B) \geq f(A \cup B) + f(A \cap B)$ definition of submodularity, submodular concave?

Submodular Concave

- Why do we call the $f(A) + f(B) \geq f(A \cup B) + f(A \cap B)$ definition of submodularity, submodular concave?
- A continuous twice differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is concave iff $\nabla^2 f \preceq 0$ (the Hessian matrix is nonpositive definite).

Submodular Concave

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

$$(\nabla_B f)(A) \triangleq f(A \cup B) - f(A \setminus B) = f\big(B|(A \setminus B)\big) \tag{4.123}$$

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 $f : \mathbb{R}^n \to \mathbb{R}$ is concave iff $\nabla^2 f \preceq 0$ (the Hessian matrix is nonpositive definite).
- Define a "discrete derivative" or difference operator defined on discrete functions $f : 2^V \to \mathbb{R}$ as follows:

$$(\nabla_B f)(A) \triangleq f(A \cup B) - f(A \setminus B) = f\big(B|(A \setminus B)\big) \tag{4.123}$$

  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 $\nabla^2 f \preceq 0$ (the Hessian matrix is nonpositive definite).
- Define a "discrete derivative" or difference operator defined on discrete functions $f : 2^V \to \mathbb{R}$ as follows:

$$(\nabla_B f)(A) \triangleq f(A \cup B) - f(A \setminus B) = f\big(B|(A \setminus B)\big) \tag{4.123}$$

  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:

$$(\nabla_B \nabla_C f)(A) = \nabla_B [\overbrace{f(A \cup C) - f(A \setminus C)}^{(\nabla_C f)(A)}] \tag{4.124}$$

$$= (\nabla_B f)(A \cup C) - (\nabla_B f)(A \setminus C) \tag{4.125}$$

$$= f(A \cup B \cup C) - f((A \cup C) \setminus B) - f((A \setminus C) \cup B) + f((A \setminus C) \setminus B) \tag{4.126}$$

Submodular Concave

- If the second difference operator everywhere nonpositive:

$$f(A \cup B \cup C) - f((A \cup C) \setminus B) \\ - f((A \setminus C) \cup B) + f(A \setminus C \setminus B) \leq 0 \qquad (4.127)$$

Submodular Concave

- If the second difference operator everywhere nonpositive:

$$f(A \cup B \cup C) - f((A \cup C) \setminus B) \\ - f((A \setminus C) \cup B) + f(A \setminus C \setminus B) \leq 0 \tag{4.127}$$

then we have the equation:

$$f((A \cup C) \setminus B) + f((A \setminus C) \cup B) \geq f(A \cup B \cup C) + f(A \setminus C \setminus B) \tag{4.128}$$

Submodular Concave

- If the second difference operator everywhere nonpositive:

$$f(A \cup B \cup C) - f((A \cup C) \setminus B) \\ - f((A \setminus C) \cup B) + f(A \setminus C \setminus B) \leq 0 \tag{4.127}$$

then we have the equation:

$$f((A \cup C) \setminus B) + f((A \setminus C) \cup B) \geq f(A \cup B \cup C) + f(A \setminus C \setminus B) \tag{4.128}$$

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

$$f(A') + f(B') \geq f(A' \cup B') + f(A' \cap B') \tag{4.129}$$

and note that A' and B' so defined can be arbitrary.

Submodular Concave

- If the second difference operator everywhere nonpositive:

$$f(A \cup B \cup C) - f((A \cup C) \setminus B) \\ - f((A \setminus C) \cup B) + f(A \setminus C \setminus B) \leq 0 \tag{4.127}$$

then we have the equation:

$$f((A \cup C) \setminus B) + f((A \setminus C) \cup B) \geq f(A \cup B \cup C) + f(A \setminus C \setminus B) \tag{4.128}$$

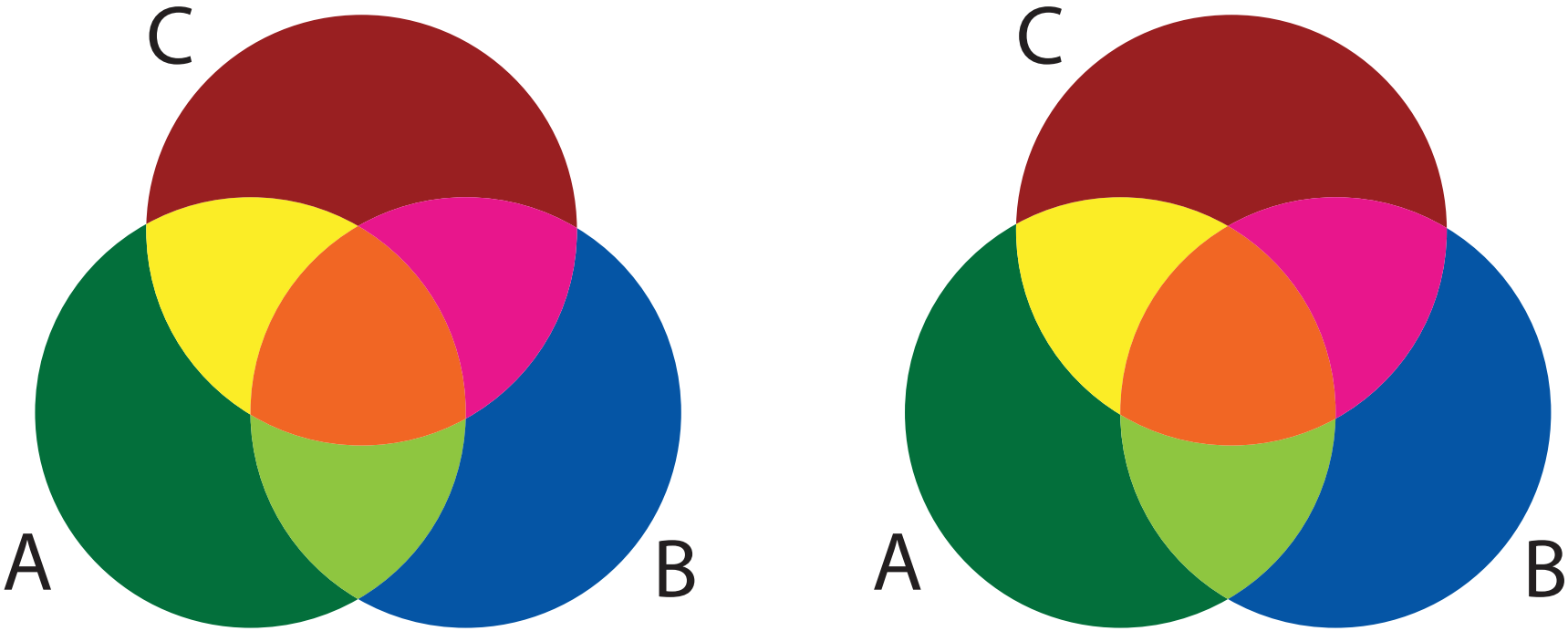
- Define $A' = (A \cup C) \setminus B$ and $B' = (A \setminus C) \cup B$. Then the above implies:

$$f(A') + f(B') \geq f(A' \cup B') + f(A' \cap B') \tag{4.129}$$

and note that A' and B' so defined can be arbitrary.

- One sense in which submodular functions are like concave functions.

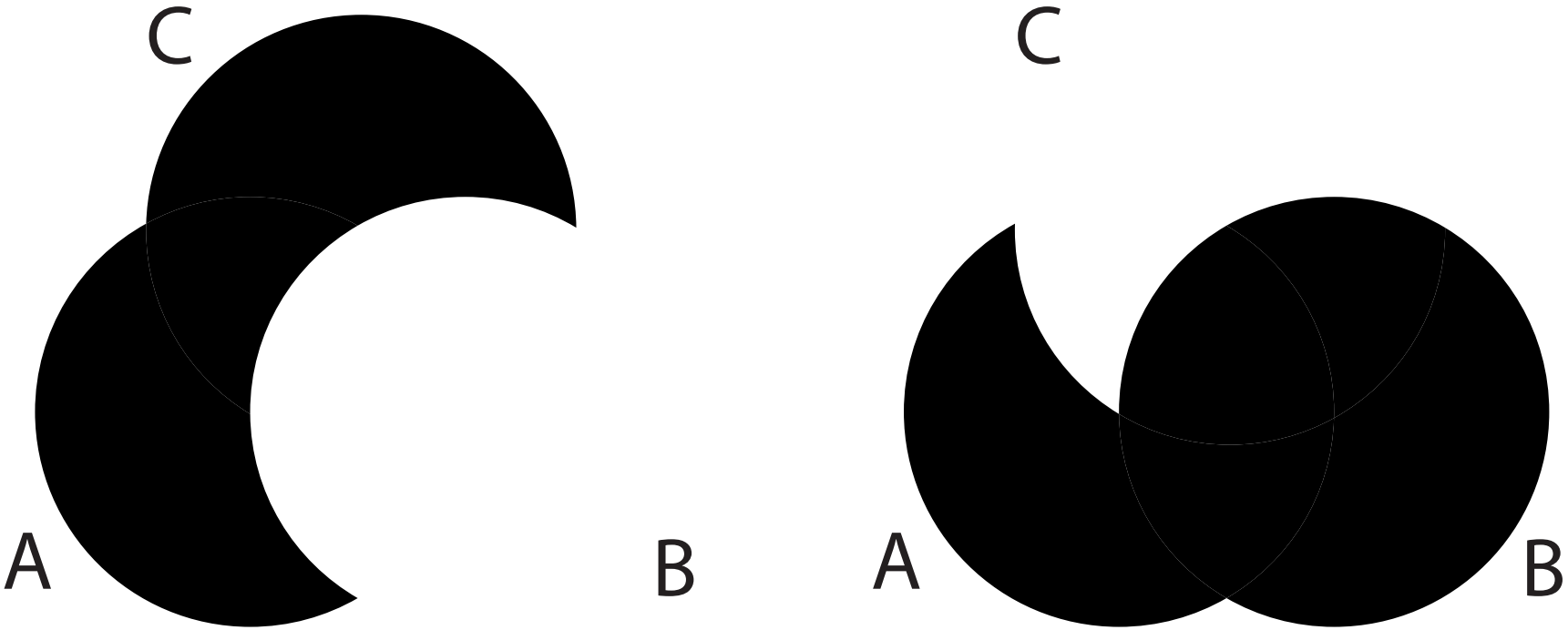
Submodular Concave



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

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

Submodular Concave



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

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

Submodularity and Concave

- This submodular/concave relationship is more simply done with singletons.

Submodularity and Concave

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

$$f(X + j) + f(X + k) \geq f(X + j + k) + f(X) \tag{4.130}$$

Submodularity and Concave

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

$$f(X + j) + f(X + k) \geq f(X + j + k) + f(X) \tag{4.130}$$

- This gives us a simpler notion corresponding to concavity.

Submodularity and Concave

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

$$f(X + j) + f(X + k) \geq f(X + j + k) + f(X) \tag{4.130}$$

- 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.

Submodularity and Concave

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

$$f(X + j) + f(X + k) \geq f(X + j + k) + f(X) \tag{4.130}$$

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

$$\nabla_j \nabla_k f(X) \leq 0 \tag{4.131}$$

Example: Rank function of a matrix

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

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

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

On Rank

- Let rank : $2^V \to \mathbb{Z}_+$ be the rank function.

On Rank

- Let rank : $2^V \to \mathbb{Z}_+$ be the rank function.
- In general, rank$(A) \leq |A|$, and vectors in A are linearly independent if and only if rank$(A) = |A|$.

On Rank

- Let rank : $2^V \to \mathbb{Z}_+$ be the rank function.
- In general, rank$(A) \leq |A|$, and vectors in A are linearly independent if and only if rank$(A) = |A|$.
- If A, B are such that rank$(A) = |A|$ and rank$(B) = |B|$, with $|A| < |B|$, then the space spanned by B is greater, and we can find a vector in B that is linearly independent of the space spanned by vectors in A.

On Rank

- Let rank : $2^V \to \mathbb{Z}_+$ be the rank function.
- In general, rank$(A) \leq |A|$, and vectors in A are linearly independent if and only if rank$(A) = |A|$.
- If A, B are such that rank$(A) = |A|$ and rank$(B) = |B|$, with $|A| < |B|$, then the space spanned by B is greater, and we can find a vector in B that is linearly independent of the space spanned by vectors in A.
- To stress this point, note that the above condition is $|A| < |B|$, **not** $A \subseteq B$ which is sufficient (to be able to find an independent vector) but not required.

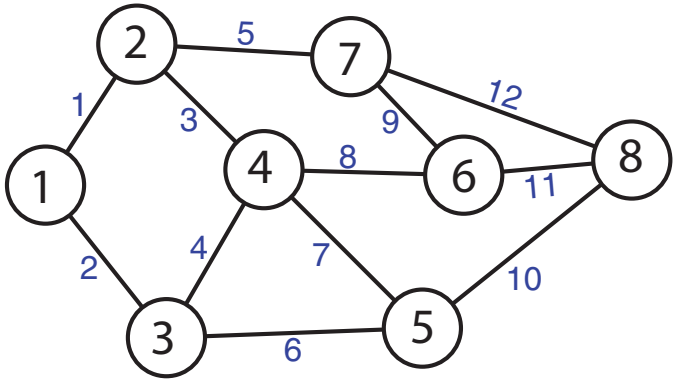
On Rank

- Let rank : $2^V \to \mathbb{Z}_+$ be the rank function.
- In general, rank$(A) \leq |A|$, and vectors in A are linearly independent if and only if rank$(A) = |A|$.
- If A, B are such that rank$(A) = |A|$ and rank$(B) = |B|$, with $|A| < |B|$, then the space spanned by B is greater, and we can find a vector in B that is linearly independent of the space spanned by vectors in A.
- To stress this point, note that the above condition is $|A| < |B|$, **not** $A \subseteq B$ which is sufficient (to be able to find an independent vector) but not required.
- In other words, given A, B with rank$(A) = |A|$ & rank$(B) = |B|$, then $|A| < |B| \Leftrightarrow \exists$ an $b \in B$ such that rank$(A \cup \{b\}) = |A| + 1$.

Spanning trees/forests

- We are given a graph $G = (V, E)$, and consider the edges $E = E(G)$ as an index set.
- Consider the $|V| \times |E|$ incidence matrix of undirected graph G, which is the matrix $\mathbf{X}_G = (x_{v,e})_{v \in V(G), e \in E(G)}$ where

$$x_{v,e} = \begin{cases} 1 & \text{if } v \in e \\ 0 & \text{if } v \notin e \end{cases} \tag{4.132}$$



$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \end{array} \begin{array}{c} \begin{array}{cccccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \end{array} \\ \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix} \end{array} \tag{4.133}$$

Spanning trees/forests & incidence matrices

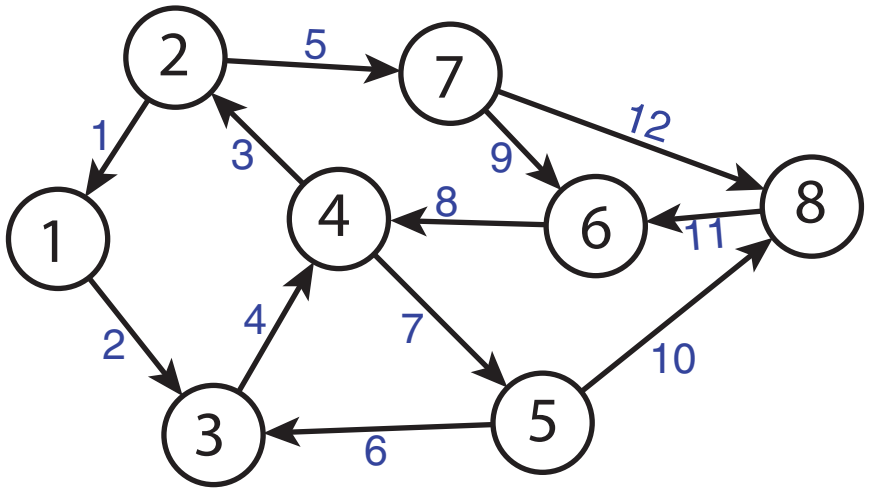
- We are given a graph $G = (V, E)$, we can arbitrarily orient the graph (make it directed) consider again the edges $E = E(G)$ as an index set.
- Consider instead the $|V| \times |E|$ incidence matrix of undirected graph G, which is the matrix $\mathbf{X}_G = (x_{v,e})_{v \in V(G), e \in E(G)}$ where

$$x_{v,e} = \begin{cases} 1 & \text{if } v \in e^+ \\ -1 & \text{if } v \in e^- \\ 0 & \text{if } v \notin e \end{cases} \tag{4.134}$$

and where e^+ is the tail and e^- is the head of (now) directed edge e.

Spanning trees/forests & incidence matrices

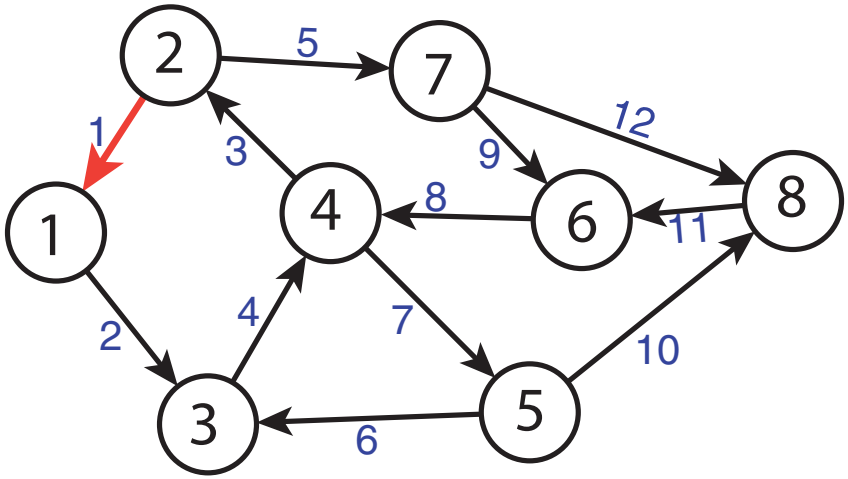
- A directed version of the graph (right) and its adjacency matrix (below).
- Orientation can be arbitrary.
- Note, rank of this matrix is 7.



$$
\begin{array}{c}
 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8
\end{array}
\begin{array}{c}
\begin{array}{cccccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \end{array} \\
\begin{pmatrix}
-1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & -1 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & -1 & 0 \\
0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & -1
\end{pmatrix}
\end{array}
$$

Spanning trees

- We can consider edge-induced subgraphs and the corresponding matrix columns.

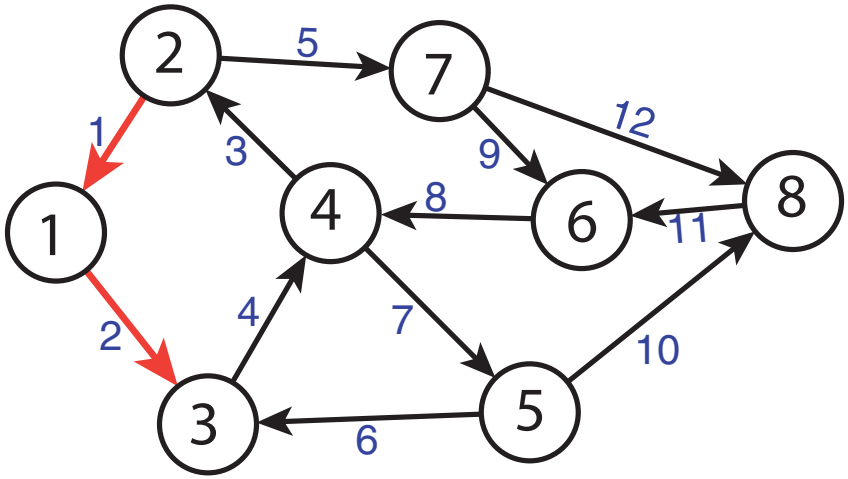


$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \end{array} \begin{array}{c} 1 \\ \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \end{array} \tag{4.135}$$

Here, $\text{rank}(\{x_1\}) = 1$.

Spanning trees

- We can consider edge-induced subgraphs and the corresponding matrix columns.

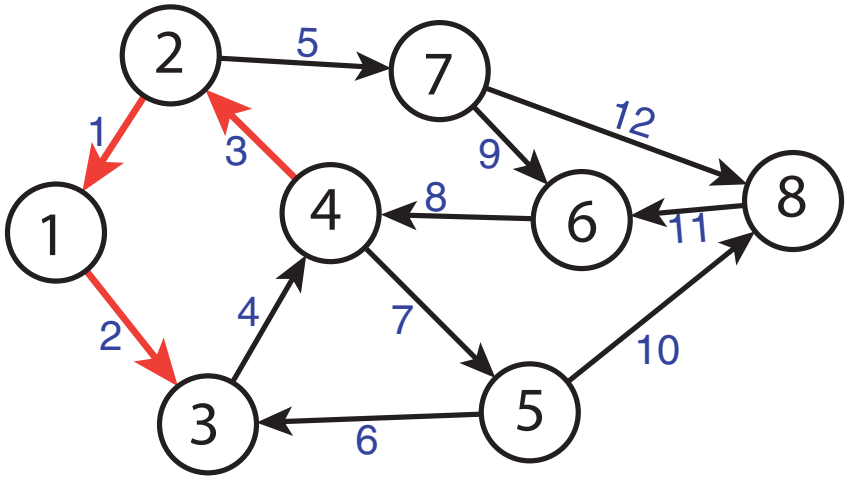


$$\begin{array}{c c} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \end{array} & \begin{pmatrix} -1 & 1 \\ 1 & 0 \\ 0 & -1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \end{array} \tag{4.135}$$

Here, rank$(\{x_1, x_2\}) = 2$.

Spanning trees

- We can consider edge-induced subgraphs and the corresponding matrix columns.

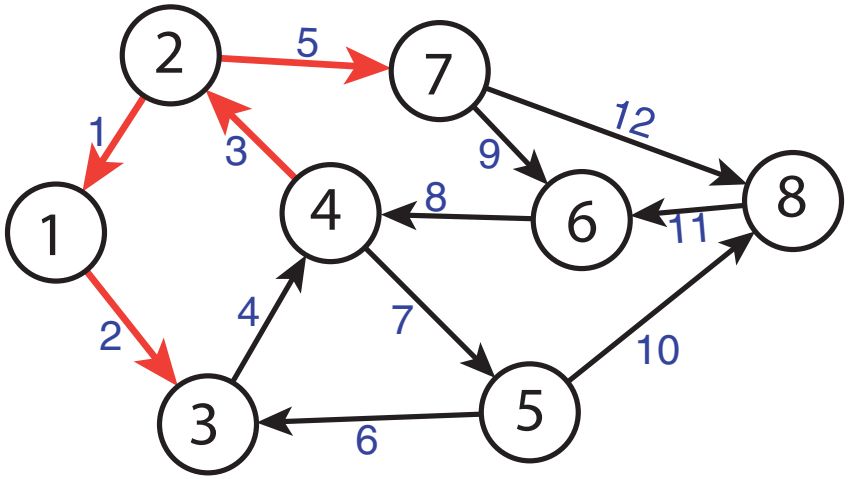


$$\begin{array}{c c} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \end{array} & \begin{pmatrix} -1 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \end{array} \tag{4.135}$$

Here, $\text{rank}(\{x_1, x_2, x_3\}) = 3$.

Spanning trees

- We can consider edge-induced subgraphs and the corresponding matrix columns.

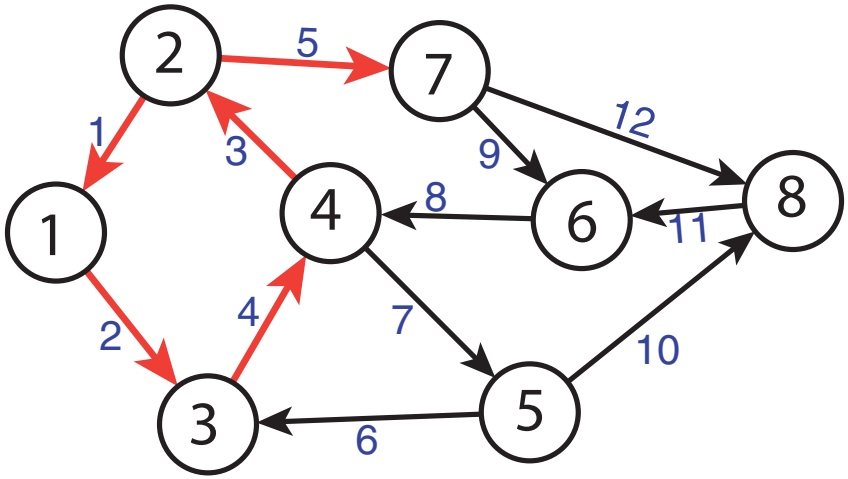


$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 5 \end{array} \\ \begin{pmatrix} -1 & 1 & 0 & 0 \\ 1 & 0 & -1 & 1 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \end{array} \tag{4.135}$$

Here, rank$(\{x_1, x_2, x_3, x_5\}) = 4$.

Spanning trees

- We can consider edge-induced subgraphs and the corresponding matrix columns.

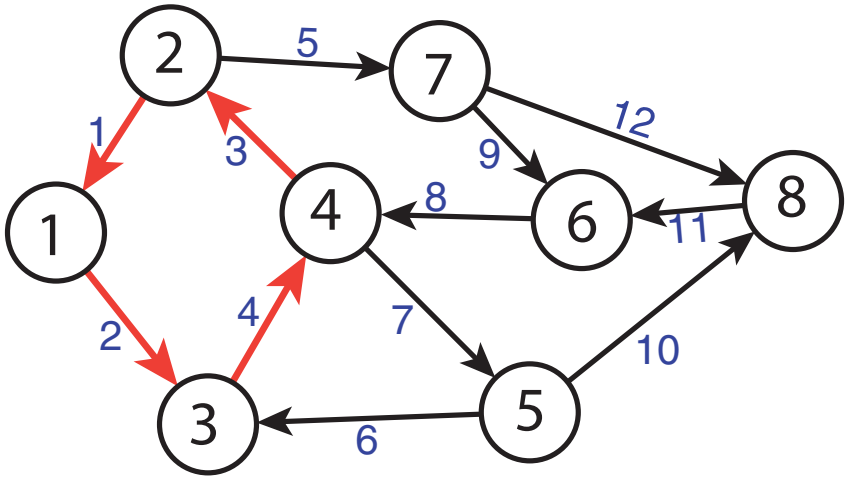


$$
\begin{array}{c}
 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8
\end{array}
\begin{array}{c}
\begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\
\begin{pmatrix}
-1 & 1 & 0 & 0 & 0 \\
1 & 0 & -1 & 0 & 1 \\
0 & -1 & 0 & 1 & 0 \\
0 & 0 & 1 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -1 \\
0 & 0 & 0 & 0 & 0
\end{pmatrix}
\end{array}
\tag{4.135}
$$

Here, rank$(\{x_1, x_2, x_3, x_4, x_5\}) = 4$.

Spanning trees

- We can consider edge-induced subgraphs and the corresponding matrix columns.



$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} -1 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \end{array} \tag{4.135}$$

Here, rank$(\{x_1, x_2, x_3, x_4\}) = 3$ since $x_4 = -x_1 - x_2 - x_3$.

Spanning trees, rank, and connected components

- In general, whenever the edges specify a cycle, there will be a linear dependence between the corresponding set of vectors in the matrix.

Spanning trees, rank, and connected components

- In general, whenever the edges specify a cycle, there will be a linear dependence between the corresponding set of vectors in the matrix.
- This means that all forests in the graph correspond to a set of linearly independent column vectors in the matrix.

Spanning trees, rank, and connected components

- In general, whenever the edges specify a cycle, there will be a linear dependence between the corresponding set of vectors in the matrix.
- This means that all forests in the graph correspond to a set of linearly independent column vectors in the matrix.
- Consider a "rank" function defined as follows: given a set of edges $A \subseteq E(G)$, the rank(A) is the size of the largest forest in the A-edge induced subgraph of G.

Spanning trees, rank, and connected components

- In general, whenever the edges specify a cycle, there will be a linear dependence between the corresponding set of vectors in the matrix.
- This means that all forests in the graph correspond to a set of linearly independent column vectors in the matrix.
- Consider a "rank" function defined as follows: given a set of edges $A \subseteq E(G)$, the rank(A) is the size of the largest forest in the A-edge induced subgraph of G.
- The rank of the entire graph then is then a spanning forest of the graph (spanning tree if the graph is connected).

Spanning trees, rank, and connected components

- In general, whenever the edges specify a cycle, there will be a linear dependence between the corresponding set of vectors in the matrix.
- This means that all forests in the graph correspond to a set of linearly independent column vectors in the matrix.
- Consider a "rank" function defined as follows: given a set of edges $A \subseteq E(G)$, the rank(A) is the size of the largest forest in the A-edge induced subgraph of G.
- The rank of the entire graph then is then a spanning forest of the graph (spanning tree if the graph is connected).
- The rank of the graph is rank$(E(G)) = |V| - k$ where k is the number of connected components of G.

Spanning trees, rank, and connected components

- In general, whenever the edges specify a cycle, there will be a linear dependence between the corresponding set of vectors in the matrix.
- This means that all forests in the graph correspond to a set of linearly independent column vectors in the matrix.
- Consider a "rank" function defined as follows: given a set of edges $A \subseteq E(G)$, the rank(A) is the size of the largest forest in the A-edge induced subgraph of G.
- The rank of the entire graph then is then a spanning forest of the graph (spanning tree if the graph is connected).
- The rank of the graph is rank$(E(G)) = |V| - k$ where k is the number of connected components of G.
- For $A \subseteq E(G)$, define $k_G(A)$ as the number of connected components of the edge-induced spanning subgraph $(V(G), A)$. Recall, $k_G(A)$ is supermodular, so $|V(G)| - k_G(A)$ is submodular.

Spanning trees, rank, and connected components

- In general, whenever the edges specify a cycle, there will be a linear dependence between the corresponding set of vectors in the matrix.
- This means that all forests in the graph correspond to a set of linearly independent column vectors in the matrix.
- Consider a "rank" function defined as follows: given a set of edges $A \subseteq E(G)$, the rank(A) is the size of the largest forest in the A-edge induced subgraph of G.
- The rank of the entire graph then is then a spanning forest of the graph (spanning tree if the graph is connected).
- The rank of the graph is rank$(E(G)) = |V| - k$ where k is the number of connected components of G.
- For $A \subseteq E(G)$, define $k_G(A)$ as the number of connected components of the edge-induced spanning subgraph $(V(G), A)$. Recall, $k_G(A)$ is supermodular, so $|V(G)| - k_G(A)$ is submodular.
- We have rank$(A) = |V(G)| - k_G(A)$.

Spanning Tree Algorithms

- We are now given a positive edge-weighted connected graph $G = (V, E, w)$ where $w : E \to \mathbb{R}_+$ is a modular function the edges of the graph. The goal is to find the minimum spanning tree (MST) of the graph.
- Given a tree T, the cost of the tree is $\text{cost}(T) = \sum_{e \in T} w(e)$, the sum of the weights of the edges.
- There are several algorithms for MST:

Algorithm 1: Kruskal's Algorithm

```
Sort the edges so that w(e_1) ≤ w(e_2) ≤ ··· ≤ w(e_m) ;
T ← (V(G), ∅) = (V, ∅) ;
for i = 1 to m do
    if E(T) ∪ {e_i} does not create a cycle in T then
        E(T) ← E(T) ∪ {e_i} ;
```

Spanning Tree Algorithms

- We are now given a positive edge-weighted connected graph $G = (V, E, w)$ where $w : E \to \mathbb{R}_+$ is a modular function the edges of the graph. The goal is to find the minimum spanning tree (MST) of the graph.
- Given a tree T, the cost of the tree is $\text{cost}(T) = \sum_{e \in T} w(e)$, the sum of the weights of the edges.
- There are several algorithms for MST:

Algorithm 2: Jarník/Prim/Dijkstra Algorithm

$T \leftarrow \emptyset$;
while T *is not a spanning tree* **do**
  $T \leftarrow T \cup \{e\}$ for $e =$ the minimum weight edge extending the tree T to a not-yet connected vertex ;

Spanning Tree Algorithms

- We are now given a positive edge-weighted connected graph $G = (V, E, w)$ where $w : E \to \mathbb{R}_+$ is a modular function the edges of the graph. The goal is to find the minimum spanning tree (MST) of the graph.
- Given a tree T, the cost of the tree is $\text{cost}(T) = \sum_{e \in T} w(e)$, the sum of the weights of the edges.
- There are several algorithms for MST:

Algorithm 3: Borůvka's Algorithm

```
F ← ∅ /* We build up the edges of a forest in F */
while G(V, F) is disconnected do
    forall components C_i of F do
        F ← F ∪ {e_i} for e_i = the min-weight edge out of C_i;
```

Spanning Tree Algorithms

- We are now given a positive edge-weighted connected graph $G = (V, E, w)$ where $w : E \to \mathbb{R}_+$ is a modular function the edges of the graph. The goal is to find the minimum spanning tree (MST) of the graph.
- Given a tree T, the cost of the tree is $\text{cost}(T) = \sum_{e \in T} w(e)$, the sum of the weights of the edges.
- There are several algorithms for MST:
- These three algorithms are all guaranteed to find the optimal minimum spanning tree in (low order) polynomial time.

Spanning Tree Algorithms

- We are now given a positive edge-weighted connected graph $G = (V, E, w)$ where $w : E \to \mathbb{R}_+$ is a modular function the edges of the graph. The goal is to find the minimum spanning tree (MST) of the graph.
- Given a tree T, the cost of the tree is $\text{cost}(T) = \sum_{e \in T} w(e)$, the sum of the weights of the edges.
- There are several algorithms for MST:
- These three algorithms are all guaranteed to find the optimal minimum spanning tree in (low order) polynomial time.
- These algorithms are all related to the "greedy" algorithm. I.e., "add next whatever looks best".

Spanning Tree Algorithms

- We are now given a positive edge-weighted connected graph $G = (V, E, w)$ where $w : E \to \mathbb{R}_+$ is a modular function the edges of the graph. The goal is to find the minimum spanning tree (MST) of the graph.
- Given a tree T, the cost of the tree is cost$(T) = \sum_{e \in T} w(e)$, the sum of the weights of the edges.
- There are several algorithms for MST:
- These three algorithms are all guaranteed to find the optimal minimum spanning tree in (low order) polynomial time.
- These algorithms are all related to the "greedy" algorithm. I.e., "add next whatever looks best".
- These algorithms will also always find a basis (a set of linearly independent vectors that span the underlying space) in the matrix example we saw earlier.

Spanning Tree Algorithms

- We are now given a positive edge-weighted connected graph $G = (V, E, w)$ where $w : E \to \mathbb{R}_+$ is a modular function the edges of the graph. The goal is to find the minimum spanning tree (MST) of the graph.
- Given a tree T, the cost of the tree is cost$(T) = \sum_{e \in T} w(e)$, the sum of the weights of the edges.
- There are several algorithms for MST:
- These three algorithms are all guaranteed to find the optimal minimum spanning tree in (low order) polynomial time.
- These algorithms are all related to the "greedy" algorithm. I.e., "add next whatever looks best".
- These algorithms will also always find a basis (a set of linearly independent vectors that span the underlying space) in the matrix example we saw earlier.
- The above are all examples of a matroid, which is the fundamental reason why the greedy algorithms work.