Submodular Functions, Optimization, and Applications to Machine Learning

— Spring Quarter, Lecture 3 —

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

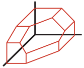
Prof. Jeff Bilmes

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

Apr 4th, 2016

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

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



Cumulative Outstanding Reading

- Read chapter 1 from Fujishige's book.

Announcements, Assignments, and Reminders

- Homework 1 is now available at our assignment dropbox (`https://canvas.uw.edu/courses/1039754/assignments`), due (electronically) Friday at 5:00pm.
- Weekly Office Hours: Mondays, 3:30-4:30, or by skype or google hangout (set up meeting via our our discussion board (`https://canvas.uw.edu/courses/1039754/discussion_topics`)).

Class Road Map - IT-I

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

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

Two Equivalent Submodular Definitions

Definition 3.2.1 (submodular concave)

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

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B) \tag{3.8}$$

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

Definition 3.2.2 (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:

$$f(A \cup \{v\}) - f(A) \geq f(B \cup \{v\}) - f(B) \tag{3.9}$$

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

Two Equivalent Supermodular Definitions

Definition 3.2.1 (supermodular)

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

$$f(A) + f(B) \leq f(A \cup B) + f(A \cap B) \tag{3.8}$$

Definition 3.2.2 (supermodular (improving returns))

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

$$f(A \cup \{v\}) - f(A) \leq f(B \cup \{v\}) - f(B) \tag{3.9}$$

- Incremental "value", "gain", or "cost" of v increases (improves) as the context in which v is considered grows from A to B.
- A function f is submodular iff $-f$ is supermodular.
- If f both submodular and supermodular, then f is said to be **modular**, and $f(A) = c + \sum_{a \in A} f(a)$ (often $c = 0$).

Submodularity's utility in ML

- A model of a physical process:
  - When maximizing, submodularity naturally models: diversity, coverage, span, and information.
  - When minimizing, submodularity naturally models: cooperative costs, complexity, roughness, and irregularity.
  - vice-versa for supermodularity.
- A submodular function can act as a parameter for a machine learning strategy (active/semi-supervised learning, discrete divergence, structured sparse convex norms for use in regularization).
- Itself, as an object or function to learn, based on data.
- A surrogate or relaxation strategy for optimization or analysis
  - An alternate to factorization, decomposition, or sum-product based simplification (as one typically finds in a graphical model). I.e., a means towards tractable surrogates for graphical models.
  - Also, we can "relax" a problem to a submodular one where it can be efficiently solved and offer a bounded quality solution.
  - Non-submodular problems can be analyzed via submodularity.

Ground set: E or V?

Submodular functions are functions defined on subsets of some finite set, called the ground set .

- It is common in the literature to use either E or V as the ground set — we will at different times use both (there should be no confusion).

Ground set: E or V?

Submodular functions are functions defined on subsets of some finite set, called the ground set .

- It is common in the literature to use either E or V as the ground set — we will at different times use both (there should be no confusion).
- The terminology ground set comes from lattice theory, where V are the ground elements of a lattice (just above 0).

Notation $\mathbb{R}^E$

What does $x \in \mathbb{R}^E$ mean?

$$\mathbb{R}^E = \{x = (x_j \in \mathbb{R} : j \in E)\} \tag{3.1}$$

$$\mathbb{R}_+^E = \{x = (x_j : j \in E) : x \geq 0\} \tag{3.2}$$

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

$$x(A) = \sum_{a \in A} x_a \tag{3.3}$$

Note that x is said to be normalized since $x(\emptyset) = 0$.

characteristic vectors of sets & modular functions

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

$$\mathbf{1}_A(j) = \begin{cases} 1 & \text{if } j \in A; \\ 0 & \text{if } j \notin A \end{cases} \tag{3.4}$$

characteristic vectors of sets & modular functions

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

$$\mathbf{1}_A(j) = \begin{cases} 1 & \text{if } j \in A; \\ 0 & \text{if } j \notin A \end{cases} \tag{3.4}$$

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

characteristic vectors of sets & modular functions

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

$$\mathbf{1}_A(j) = \begin{cases} 1 & \text{if } j \in A; \\ 0 & \text{if } j \notin A \end{cases} \tag{3.4}$$

- Sometimes this will be written as $\chi_A \equiv \mathbf{1}_A$.
- Thus, given modular function $x \in \mathbb{R}^E$, we can write $x(A)$ in a variety of ways, i.e.,

$$x(A) = x \cdot \mathbf{1}_A = \sum_{i \in A} x(i) \tag{3.5}$$

Other Notation: singletons and sets

When A is a set and k is a singleton (i.e., a single item), the union is properly written as $A \cup \{k\}$, but sometimes we will write just $A + k$.

What does S^T mean when S and T are arbitrary sets?

- Let S and T be two arbitrary sets (either of which could be countable, or uncountable).

What does S^T mean when S and T are arbitrary sets?

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

What does S^T mean when S and T are arbitrary sets?

- Let S and T be two arbitrary sets (either of which could be countable, or uncountable).
- We define the notation S^T to be the set of all functions that map from T to S. That is, if $f \in S^T$, then $f : T \to S$.
- Hence, given a finite set E, $\mathbb{R}^E$ is the set of all functions that map from elements of E to the reals $\mathbb{R}$, and such functions are identical to a vector in a vector space with axes labeled as elements of E (i.e., if $m \in \mathbb{R}^E$, then for all $e \in E$, $m(e) \in \mathbb{R}$).

What does S^T mean when S and T are arbitrary sets?

- Let S and T be two arbitrary sets (either of which could be countable, or uncountable).
- We define the notation S^T to be the set of all functions that map from T to S. That is, if $f \in S^T$, then $f : T \to S$.
- Hence, given a finite set E, $\mathbb{R}^E$ is the set of all functions that map from elements of E to the reals $\mathbb{R}$, and such functions are identical to a vector in a vector space with axes labeled as elements of E (i.e., if $m \in \mathbb{R}^E$, then for all $e \in E$, $m(e) \in \mathbb{R}$).
- Often "2" is shorthand for the set $\{0, 1\}$. I.e., $\mathbb{R}^2$ where $2 \equiv \{0, 1\}$.

What does S^T mean when S and T are arbitrary sets?

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

What does S^T mean when S and T are arbitrary sets?

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

What does S^T mean when S and T are arbitrary sets?

- Let S and T be two arbitrary sets (either of which could be countable, or uncountable).
- We define the notation S^T to be the set of all functions that map from T to S. That is, if $f \in S^T$, then $f : T \to S$.
- Hence, given a finite set E, $\mathbb{R}^E$ is the set of all functions that map from elements of E to the reals $\mathbb{R}$, and such functions are identical to a vector in a vector space with axes labeled as elements of E (i.e., if $m \in \mathbb{R}^E$, then for all $e \in E$, $m(e) \in \mathbb{R}$).
- Often "2" is shorthand for the set $\{0, 1\}$. I.e., $\mathbb{R}^2$ where $2 \equiv \{0, 1\}$.
- Similarly, 2^E is the set of all functions from E to "two" — so 2^E is shorthand for $\{0, 1\}^E$ — hence, 2^E is the set of all functions that map from elements of E to $\{0, 1\}$, equivalent to all binary vectors with elements indexed by elements of E, equivalent to subsets of E. Hence, if $A \in 2^E$ then $A \subseteq E$.
- What might 3^E mean?

Example Submodular: Entropy from Information Theory

- Entropy is submodular. Let V be the index set of a set of random variables, then the function

$$f(A) = H(X_A) = -\sum_{x_A} p(x_A) \log p(x_A) \tag{3.6}$$

is submodular.

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

$$\begin{aligned} H(X_v|X_B) &= H(X_{B+v}) - H(X_B) && (3.7)\\ &\leq H(X_{A+v}) - H(X_A) = H(X_v|X_A) && (3.8)\end{aligned}$$

- We say "further" due to $B \setminus A$ not nec. empty.

Example Submodular: Entropy from Information Theory

- Alternate Proof: Conditional mutual Information is always non-negative.
- Given $A, B \subseteq V$, consider conditional mutual information quantity:

$$I(X_{A \setminus B}; X_{B \setminus A} | X_{A \cap B}) = \sum_{x_{A \cup B}} p(x_{A \cup B}) \log \frac{p(x_{A \setminus B}, x_{B \setminus A} | x_{A \cap B})}{p(x_{A \setminus B} | x_{A \cap B}) p(x_{B \setminus A} | x_{A \cap B})}$$

$$= \sum_{x_{A \cup B}} p(x_{A \cup B}) \log \frac{p(x_{A \cup B}) p(x_{A \cap B})}{p(x_A) p(x_B)} \geq 0 \quad (3.9)$$

then

$$I(X_{A \setminus B}; X_{B \setminus A} | X_{A \cap B})$$
$$= H(X_A) + H(X_B) - H(X_{A \cup B}) - H(X_{A \cap B}) \geq 0 \quad (3.10)$$

so entropy satisfies

$$H(X_A) + H(X_B) \geq H(X_{A \cup B}) + H(X_{A \cap B}) \quad (3.11)$$

Information Theory: Block Coding

- Given a set of random variables $\{X_i\}_{i \in V}$ indexed by set V, how do we partition them so that we can best block-code them within each block.

Information Theory: Block Coding

- Given a set of random variables $\{X_i\}_{i \in V}$ indexed by set V, how do we partition them so that we can best block-code them within each block.
- I.e., how do we form $S \subseteq V$ such that $I(X_S; X_{V \setminus S})$ is as small as possible, where $I(X_A; X_B)$ is the mutual information between random variables X_A and X_B, i.e.,

$$I(X_A; X_B) = H(X_A) + H(X_B) - H(X_A, X_B) \tag{3.12}$$

and $H(X_A) = -\sum_{x_A} p(x_A) \log p(x_A)$ is the joint entropy of the set X_A of random variables.

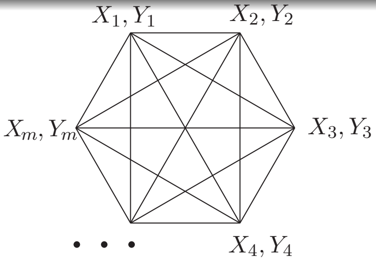
Example Submodular: Mutual Information

- Also, symmetric mutual information is submodular,

$$f(A) = I(X_A; X_{V \setminus A}) = H(X_A) + H(X_{V \setminus A}) - H(X_V) \tag{3.13}$$

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

Information Theory: Network Communication



- A network of senders/receivers
- Each sender X_i is trying to communicate simultaneously with each receiver Y_i (i.e., for all i, X_i is sending to $\{Y_i\}_i$
- The X_i are not necessarily independent.

- Communication rates from i to j are $R^{(i \to j)}$ to send message $W^{(i \to j)} \in \left\{1, 2, \ldots, 2^{nR^{(i \to j)}}\right\}$.
- Goal: necessary and sufficient conditions for achievability.
- I.e., can we find functions f such that any rates must satisfy

$$\forall S \subseteq V, \quad \sum_{i \in S, j \in V \setminus S} R^{(i \to j)} \leq f(S) \tag{3.14}$$

- Special cases MAC (Multi-Access Channel) for communication over $p(y|x_1, x_2)$ and Slepian-Wolf compression (independent compression of X and Y but at joint rate $H(X, Y)$).

Monge Matrices

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

$$c_{ij} + c_{rs} \leq c_{is} + c_{rj} \tag{3.15}$$

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

Monge Matrices

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

$$c_{ij} + c_{rs} \leq c_{is} + c_{rj} \tag{3.15}$$

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

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

$$c_{\min(i,r),\min(j,s)} + c_{\max(i,r),\max(j,s)} \leq c_{is} + c_{rj} \tag{3.16}$$

Monge Matrices

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

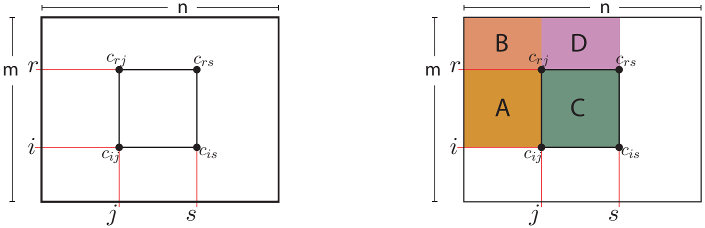
$$c_{ij} + c_{rs} \leq c_{is} + c_{rj} \tag{3.15}$$

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

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

$$c_{\min(i,r),\min(j,s)} + c_{\max(i,r),\max(j,s)} \leq c_{is} + c_{rj} \tag{3.16}$$

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



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

Monge Matrices, where useful

- Useful for speeding up many transportation, dynamic programming, flow, search, lot-sizing and many other problems.

Monge Matrices, where useful

- Useful for speeding up many transportation, dynamic programming, flow, search, lot-sizing and many other problems.
- Example, Hitchcock transportation problem: Given $m \times n$ cost matrix $C = [c_{ij}]_{ij}$, a non-negative supply vector $a \in \mathbb{R}^m_+$, a non-negative demand vector $b \in \mathbb{R}^n_+$ with $\sum_{i=1}^m a(i) = \sum_{j=1}^n b_j$, we wish to optimally solve the following linear program:

$$\underset{X \in \mathbb{R}^{m \times n}}{\text{minimize}} \quad \sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij} \tag{3.17}$$

$$\text{subject to} \quad \sum_{i=1}^m x_{ij} = b_j \;\; \forall j = 1, \dots, n \tag{3.18}$$

$$\sum_{j=1}^n x_{ij} = a_i \;\; \forall i = 1, \dots, m \tag{3.19}$$

$$x_{i,j} \geq 0 \;\; \forall i, j \tag{3.20}$$

Monge Matrices, Hitchcock transportation

C

Producers, Sources, or Supply

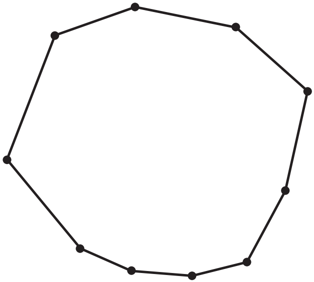
		b_1	b_2	b_3	b_4
a_1	2	0	1	3	3
a_2	1	1	4	7	10
a_3	5	0	4	9	14
		3	2	1	2

Consumers, Sinks, or Demand

- Solving the linear program can be done easily and optimally using the "North West Corner Rule" in only $O(m+n)$ if the matrix C is Monge!

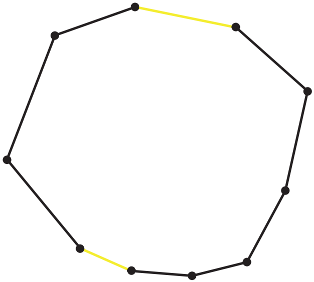
Monge Matrices and Convex Polygons

- Can generate a Monge matrix from a convex polygon - delete two segments, then separately number vertices on each chain. Distances c_{ij} satisfy Monge property (or quadrangle inequality).



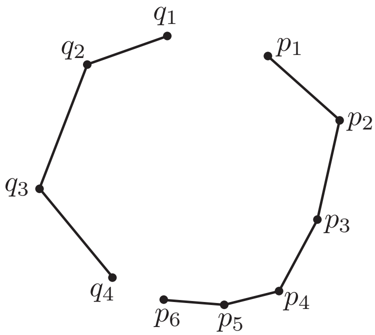
Monge Matrices and Convex Polygons

- Can generate a Monge matrix from a convex polygon - delete two segments, then separately number vertices on each chain. Distances c_{ij} satisfy Monge property (or quadrangle inequality).



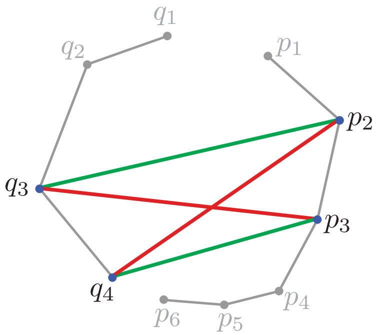
Monge Matrices and Convex Polygons

- Can generate a Monge matrix from a convex polygon - delete two segments, then separately number vertices on each chain. Distances c_{ij} satisfy Monge property (or quadrangle inequality).



Monge Matrices and Convex Polygons

- Can generate a Monge matrix from a convex polygon - delete two segments, then separately number vertices on each chain. Distances c_{ij} satisfy Monge property (or quadrangle inequality).



$$d(p_2, q_3) + d(p_3, q_4) \leq d(p_2, q_4) + d(p_3, q_3) \tag{3.21}$$

Monge Matrices and Submodularity

- A submodular function has the form: $f : 2^V \to \mathbb{R}$ which can be seen as $f : \{0,1\}^V \to \mathbb{R}$

Monge Matrices and Submodularity

- A submodular function has the form: $f : 2^V \to \mathbb{R}$ which can be seen as $f : \{0,1\}^V \to \mathbb{R}$
- We can generalize this to $f : \{0,K\}^V \to \mathbb{R}$ for some constant $K \in \mathbb{Z}_+$.

Monge Matrices and Submodularity

- A submodular function has the form: $f : 2^V \to \mathbb{R}$ which can be seen as $f : \{0,1\}^V \to \mathbb{R}$
- We can generalize this to $f : \{0,K\}^V \to \mathbb{R}$ for some constant $K \in \mathbb{Z}_+$.
- We may define submodularity as: for all $x, y \in \{0,K\}^V$, we have

$$f(x) + f(y) \geq f(x \vee y) + f(x \wedge y) \tag{3.22}$$

Monge Matrices and Submodularity

- A submodular function has the form: $f : 2^V \to \mathbb{R}$ which can be seen as $f : \{0,1\}^V \to \mathbb{R}$
- We can generalize this to $f : \{0,K\}^V \to \mathbb{R}$ for some constant $K \in \mathbb{Z}_+$.
- We may define submodularity as: for all $x, y \in \{0,K\}^V$, we have

$$f(x) + f(y) \geq f(x \vee y) + f(x \wedge y) \tag{3.22}$$

- $x \vee y$ is the (join) element-wise min of each element, that is $(x \vee y)(v) = \min(x(v), y(v))$ for $v \in V$.

Monge Matrices and Submodularity

- A submodular function has the form: $f : 2^V \to \mathbb{R}$ which can be seen as $f : \{0, 1\}^V \to \mathbb{R}$
- We can generalize this to $f : \{0, K\}^V \to \mathbb{R}$ for some constant $K \in \mathbb{Z}_+$.
- We may define submodularity as: for all $x, y \in \{0, K\}^V$, we have

$$f(x) + f(y) \geq f(x \vee y) + f(x \wedge y) \tag{3.22}$$

- $x \vee y$ is the (join) element-wise min of each element, that is $(x \vee y)(v) = \min(x(v), y(v))$ for $v \in V$.
- $x \wedge y$ is the (meet) element-wise min of each element, that is, $(x \wedge y)(v) = \max(x(v), y(v))$ for $v \in V$.

Monge Matrices and Submodularity

- A submodular function has the form: $f : 2^V \to \mathbb{R}$ which can be seen as $f : \{0,1\}^V \to \mathbb{R}$
- We can generalize this to $f : \{0,K\}^V \to \mathbb{R}$ for some constant $K \in \mathbb{Z}_+$.
- We may define submodularity as: for all $x, y \in \{0,K\}^V$, we have

$$f(x) + f(y) \geq f(x \vee y) + f(x \wedge y) \tag{3.22}$$

- $x \vee y$ is the (join) element-wise min of each element, that is $(x \vee y)(v) = \min(x(v), y(v))$ for $v \in V$.
- $x \wedge y$ is the (meet) element-wise min of each element, that is, $(x \wedge y)(v) = \max(x(v), y(v))$ for $v \in V$.
- With $K = 1$, then this is the standard definition of submodularity.

Monge Matrices and Submodularity

- A submodular function has the form: $f : 2^V \to \mathbb{R}$ which can be seen as $f : \{0,1\}^V \to \mathbb{R}$
- We can generalize this to $f : \{0,K\}^V \to \mathbb{R}$ for some constant $K \in \mathbb{Z}_+$.
- We may define submodularity as: for all $x, y \in \{0,K\}^V$, we have

$$f(x) + f(y) \geq f(x \vee y) + f(x \wedge y) \tag{3.22}$$

- $x \vee y$ is the (join) element-wise min of each element, that is $(x \vee y)(v) = \min(x(v), y(v))$ for $v \in V$.
- $x \wedge y$ is the (meet) element-wise min of each element, that is, $(x \wedge y)(v) = \max(x(v), y(v))$ for $v \in V$.
- With $K = 1$, then this is the standard definition of submodularity.
- With $|V| = 2$, and $K + 1$ the side-dimension of the matrix, we get a Monge property (on square matrices).

Two Equivalent Submodular Definitions

Definition 3.6.1 (submodular concave)

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

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B) \tag{3.8}$$

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

Definition 3.6.2 (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:

$$f(A \cup \{v\}) - f(A) \geq f(B \cup \{v\}) - f(B) \tag{3.9}$$

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

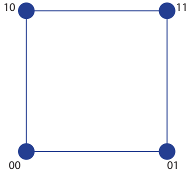
Submodular on Hypercube Verticies

- Test submodularity via values on verticies of hypercube.

Submodular on Hypercube Verticies

- Test submodularity via values on verticies of hypercube.

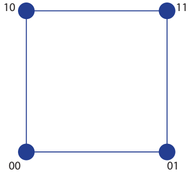
Example: with $|V| = n = 2$, this is easy:



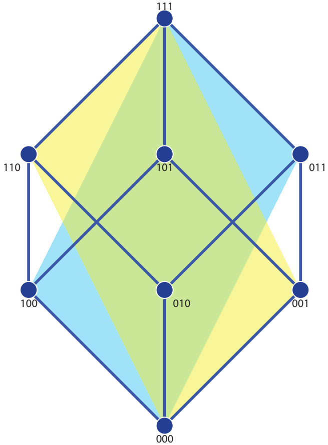
Submodular on Hypercube Verticies

- Test submodularity via values on verticies of hypercube.

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



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



How many inequalities?

Subadditive Definitions

Definition 3.6.1 (subadditive)

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

$$f(A) + f(B) \geq f(A \cup B) \tag{3.23}$$

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

Two Equivalent Supermodular Definitions

Definition 3.6.1 (supermodular)

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

$$f(A) + f(B) \leq f(A \cup B) + f(A \cap B) \tag{3.8}$$

Definition 3.6.2 (supermodular (improving returns))

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

$$f(A \cup \{v\}) - f(A) \leq f(B \cup \{v\}) - f(B) \tag{3.9}$$

- Incremental "value", "gain", or "cost" of v increases (improves) as the context in which v is considered grows from A to B.
- A function f is submodular iff $-f$ is supermodular.
- If f both submodular and supermodular, then f is said to be **modular**, and $f(A) = c + \sum_{a \in A} f(a)$ (often $c = 0$).

Superadditive Definitions

Definition 3.6.2 (superadditive)

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

$$f(A) + f(B) \leq f(A \cup B) \tag{3.24}$$

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

Superadditive Definitions

Definition 3.6.2 (superadditive)

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

$$f(A) + f(B) \leq f(A \cup B) \tag{3.24}$$

- 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.

Superadditive Definitions

Definition 3.6.2 (superadditive)

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

$$f(A) + f(B) \leq f(A \cup B) \tag{3.24}$$

- 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:

$$f(A) = \begin{cases} 1 & \text{if } |A| \leq k \\ 0 & \text{else} \end{cases} \tag{3.25}$$

Superadditive Definitions

Definition 3.6.2 (superadditive)

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

$$f(A) + f(B) \leq f(A \cup B) \tag{3.24}$$

- 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:

$$f(A) = \begin{cases} 1 & \text{if } |A| \leq k \\ 0 & \text{else} \end{cases} \tag{3.25}$$

- This function is subadditive but not submodular.

Modular Definitions

Definition 3.6.3 (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

$$f(A) + f(B) = f(A \cap B) + f(A \cup B) \tag{3.26}$$

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

Proposition 3.6.4

If f is modular, it may be written as

$$f(A) = f(\emptyset) + \sum_{a \in A} \Big(f(\{a\}) - f(\emptyset) \Big) = c + \sum_{a \in A} f'(a) \tag{3.27}$$

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

Modular Definitions

Proof.

We inductively construct the value for $A = \{a_1, a_2, \dots, a_k\}$.
For $k = 2$,

$$f(a_1) + f(a_2) = f(a_1, a_2) + f(\emptyset) \tag{3.28}$$

$$\text{implies } f(a_1, a_2) = f(a_1) - f(\emptyset) + f(a_2) - f(\emptyset) + f(\emptyset) \tag{3.29}$$

then for $k = 3$,

$$f(a_1, a_2) + f(a_3) = f(a_1, a_2, a_3) + f(\emptyset) \tag{3.30}$$

$$\text{implies } f(a_1, a_2, a_3) = f(a_1, a_2) - f(\emptyset) + f(a_3) - f(\emptyset) + f(\emptyset) \tag{3.31}$$

$$= f(\emptyset) + \sum_{i=1}^{3} \big(f(a_i) - f(\emptyset)\big) \tag{3.32}$$

and so on ... □

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 3.6.5

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

Proof.

$$\bar{f}(A) + \bar{f}(B) \geq \bar{f}(A \cup B) + \bar{f}(A \cap B) \tag{3.33}$$

follows from

$$f(V \setminus A) + f(V \setminus B) \geq f(V \setminus (A \cup B)) + f(V \setminus (A \cap B)) \tag{3.34}$$

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

Undirected Graphs

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

Undirected Graphs

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

$$E(X, Y) = \{\{x, y\} \in E(G) : x \in X \setminus Y, y \in Y \setminus X\} \tag{3.35}$$

as the edges strictly between X and Y.

Undirected Graphs

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

$$E(X, Y) = \{\{x, y\} \in E(G) : x \in X \setminus Y, y \in Y \setminus X\} \tag{3.35}$$

as the edges strictly between X and Y.
- Nodes define cuts, define the cut function $\delta(X) = E(X, V \setminus X)$.

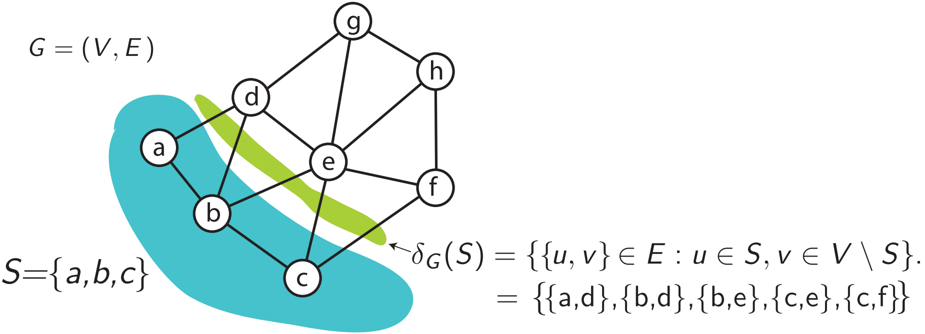
Undirected Graphs

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

$$E(X,Y) = \{\{x,y\} \in E(G) : x \in X \setminus Y, y \in Y \setminus X\} \tag{3.35}$$

as the edges strictly between X and Y.

- Nodes define cuts, define the cut function $\delta(X) = E(X, V \setminus X)$.



Directed graphs, and cuts and flows

- If G is directed, define

$$E^+(X,Y) \triangleq \{(x,y) \in E(G) : x \in X \setminus Y, y \in Y \setminus X\} \tag{3.36}$$

as the edges directed strictly from X towards Y.

Directed graphs, and cuts and flows

- If G is directed, define

$$E^+(X,Y) \triangleq \{(x,y) \in E(G) : x \in X \setminus Y, y \in Y \setminus X\} \tag{3.36}$$

as the edges directed strictly from X towards Y.

- Nodes define cuts and flows. Define edges leaving X (out-flow) as

$$\delta^+(X) \triangleq E^+(X, V \setminus X) \tag{3.37}$$

and edges entering X (in-flow) as

$$\delta^-(X) \triangleq E^+(V \setminus X, X) \tag{3.38}$$

Directed graphs, and cuts and flows

- If G is directed, define

$$E^+(X,Y) \triangleq \{(x,y) \in E(G) : x \in X \setminus Y, y \in Y \setminus X\} \tag{3.36}$$

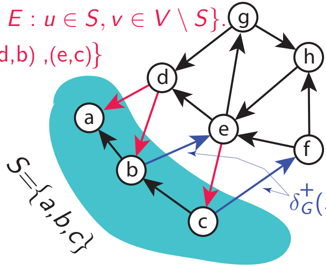
  as the edges directed strictly from X towards Y.
- Nodes define cuts and flows. Define edges leaving X (out-flow) as

$$\delta^+(X) \triangleq E^+(X, V \setminus X) \tag{3.37}$$

  and edges entering X (in-flow) as

$$\delta^-(X) \triangleq E^+(V \setminus X, X) \tag{3.38}$$

$$\delta^-_G(S) = \{(v,u) \in E : u \in S, v \in V \setminus S\}. = \{(d,a), (d,b), (e,c)\}$$



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

The Neighbor function in undirected graphs

- Given a set $X \subseteq V$, the neighbor function of X is defined as

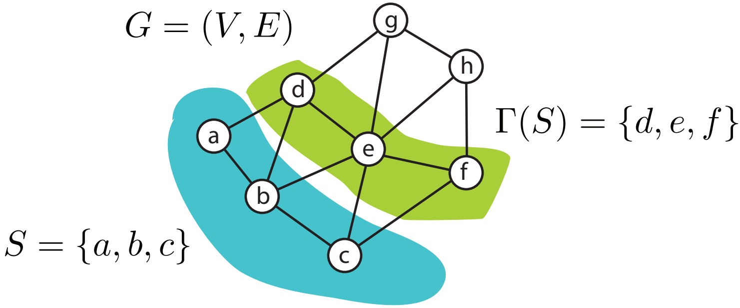
$$\Gamma(X) \triangleq \{v \in V(G) \setminus X : E(X, \{v\}) \neq \emptyset\} \tag{3.39}$$

The Neighbor function in undirected graphs

- Given a set $X \subseteq V$, the neighbor function of X is defined as

$$\Gamma(X) \triangleq \{v \in V(G) \setminus X : E(X, \{v\}) \neq \emptyset\} \tag{3.39}$$

- Example:



Directed Cut function: property

Lemma 3.7.1

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

$$|\delta^+(X)| + |\delta^+(Y)| \\ = |\delta^+(X \cap Y)| + |\delta^+(X \cup Y)| + |E^+(X, Y)| + |E^+(Y, X)| \tag{3.40}$$

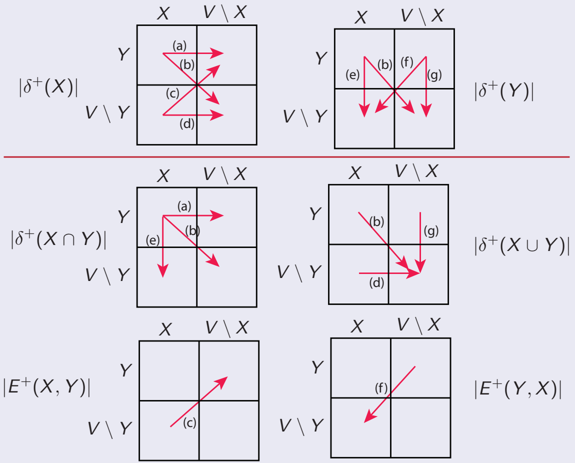
and

$$|\delta^-(X)| + |\delta^-(Y)| \\ = |\delta^-(X \cap Y)| + |\delta^-(X \cup Y)| + |E^-(X, Y)| + |E^-(Y, X)| \tag{3.41}$$

Directed Cut function: proof of property

Proof.

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



Directed cut/flow functions: submodular

Lemma 3.7.2

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

Proof.

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

More generally, in the non-negative weighted case, both in-flow and out-flow are submodular on subsets of the vertices.

Undirected Cut/Flow & the Neighbor function: submodular

Lemma 3.7.3

For an undirected graph $G = (V, E)$ and any $X, Y \subseteq V$: we have that both the undirected cut (or flow) function $|\delta(X)|$ and the neighbor function $|\Gamma(X)|$ are submodular. I.e.,

$$|\delta(X)| + |\delta(Y)| = |\delta(X \cap Y)| + |\delta(X \cup Y)| + 2|E(X, Y)| \tag{3.42}$$

and

$$|\Gamma(X)| + |\Gamma(Y)| \geq |\Gamma(X \cap Y)| + |\Gamma(X \cup Y)| \tag{3.43}$$

Proof.

- Eq. (3.42) follows from Eq. (3.40): we replace each undirected edge $\{u, v\}$ with two oppositely-directed directed edges (u, v) and (v, u). Then we use same counting argument.

. . .

Undirected Cut/Flow & the Neighbor function: submodular

Lemma 3.7.3

For an undirected graph $G = (V, E)$ and any $X, Y \subseteq V$: we have that both the undirected cut (or flow) function $|\delta(X)|$ and the neighbor function $|\Gamma(X)|$ are submodular. I.e.,

$$|\delta(X)| + |\delta(Y)| = |\delta(X \cap Y)| + |\delta(X \cup Y)| + 2|E(X, Y)| \tag{3.42}$$

and

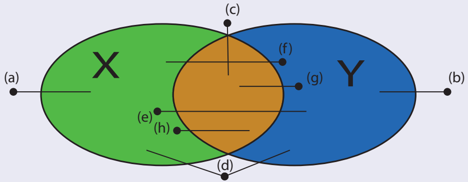
$$|\Gamma(X)| + |\Gamma(Y)| \geq |\Gamma(X \cap Y)| + |\Gamma(X \cup Y)| \tag{3.43}$$

Proof.

- Eq. (3.42) follows from Eq. (3.40): we replace each undirected edge $\{u, v\}$ with two oppositely-directed directed edges (u, v) and (v, u). Then we use same counting argument.
- Eq. (3.43) follows as shown in the following page.

. . .

cont.



Graphically, we can count and see that

$$\Gamma(X) = (a) + (c) + (f) + (g) + (d) \tag{3.44}$$
$$\Gamma(Y) = (b) + (c) + (e) + (h) + (d) \tag{3.45}$$
$$\Gamma(X \cup Y) = (a) + (b) + (c) + (d) \tag{3.46}$$
$$\Gamma(X \cap Y) = (c) + (g) + (h) \tag{3.47}$$

so

$$\begin{aligned}|\Gamma(X)| + |\Gamma(Y)| &= (a) + (b) + 2(c) + 2(d) + (e) + (f) + (g) + (h) \\ &\geq (a) + (b) + 2(c) + (d) + (g) + (h) = |\Gamma(X \cup Y)| + |\Gamma(X \cap Y)|\end{aligned} \tag{3.48}$$

Undirected Neighbor functions

Therefore, the undirected cut function $|\delta(A)|$ and the neighbor function $|\Gamma(A)|$ of a graph G are both submodular.

Undirected cut/flow is submodular: alternate proof

- Another simple proof shows that $|\delta(X)|$ is submodular.

Undirected cut/flow is submodular: alternate proof

- Another simple proof shows that $|\delta(X)|$ is submodular.
- Define a graph $G_{uv} = (\{u, v\}, \{e\}, w)$ with two nodes u, v and one edge $e = \{u, v\}$ with non-negative weight $w(e) \in \mathbb{R}_+$.

Undirected cut/flow is submodular: alternate proof

- Another simple proof shows that $|\delta(X)|$ is submodular.
- Define a graph $G_{uv} = (\{u, v\}, \{e\}, w)$ with two nodes u, v and one edge $e = \{u, v\}$ with non-negative weight $w(e) \in \mathbb{R}_+$.
- Cut weight function over those two nodes: $w(\delta_{u,v}(\cdot))$ has valuation:

$$w(\delta_{u,v}(\emptyset)) = w(\delta_{u,v}(\{u, v\})) = 0 \tag{3.49}$$

and

$$w(\delta_{u,v}(\{u\})) = w(\delta_{u,v}(\{v\})) = w \geq 0 \tag{3.50}$$

Undirected cut/flow is submodular: alternate proof

- Another simple proof shows that $|\delta(X)|$ is submodular.
- Define a graph $G_{uv} = (\{u, v\}, \{e\}, w)$ with two nodes u, v and one edge $e = \{u, v\}$ with non-negative weight $w(e) \in \mathbb{R}_+$.
- Cut weight function over those two nodes: $w(\delta_{u,v}(\cdot))$ has valuation:

$$w(\delta_{u,v}(\emptyset)) = w(\delta_{u,v}(\{u, v\})) = 0 \tag{3.49}$$

and

$$w(\delta_{u,v}(\{u\})) = w(\delta_{u,v}(\{v\})) = w \geq 0 \tag{3.50}$$

- Thus, $w(\delta_{u,v}(\cdot))$ is submodular since

$$w(\delta_{u,v}(\{u\})) + w(\delta_{u,v}(\{v\})) \geq w(\delta_{u,v}(\{u, v\})) + w(\delta_{u,v}(\emptyset)) \tag{3.51}$$

Undirected cut/flow is submodular: alternate proof

- Another simple proof shows that $|\delta(X)|$ is submodular.
- Define a graph $G_{uv} = (\{u,v\},\{e\},w)$ with two nodes u,v and one edge $e = \{u,v\}$ with non-negative weight $w(e) \in \mathbb{R}_+$.
- Cut weight function over those two nodes: $w(\delta_{u,v}(\cdot))$ has valuation:

$$w(\delta_{u,v}(\emptyset)) = w(\delta_{u,v}(\{u,v\})) = 0 \tag{3.49}$$

and

$$w(\delta_{u,v}(\{u\})) = w(\delta_{u,v}(\{v\})) = w \geq 0 \tag{3.50}$$

- Thus, $w(\delta_{u,v}(\cdot))$ is submodular since

$$w(\delta_{u,v}(\{u\})) + w(\delta_{u,v}(\{v\})) \geq w(\delta_{u,v}(\{u,v\})) + w(\delta_{u,v}(\emptyset)) \tag{3.51}$$

- General non-negative weighted graph $G = (V,E,w)$, define $w(\delta(\cdot))$:

$$f(X) = w(\delta(X)) = \sum_{(u,v)\in E(G)} w(\delta_{u,v}(X \cap \{u,v\})) \tag{3.52}$$

Undirected cut/flow is submodular: alternate proof

- Another simple proof shows that $|\delta(X)|$ is submodular.
- Define a graph $G_{uv} = (\{u, v\}, \{e\}, w)$ with two nodes u, v and one edge $e = \{u, v\}$ with non-negative weight $w(e) \in \mathbb{R}_+$.
- Cut weight function over those two nodes: $w(\delta_{u,v}(\cdot))$ has valuation:

$$w(\delta_{u,v}(\emptyset)) = w(\delta_{u,v}(\{u, v\})) = 0 \tag{3.49}$$

and

$$w(\delta_{u,v}(\{u\})) = w(\delta_{u,v}(\{v\})) = w \geq 0 \tag{3.50}$$

- Thus, $w(\delta_{u,v}(\cdot))$ is submodular since

$$w(\delta_{u,v}(\{u\})) + w(\delta_{u,v}(\{v\})) \geq w(\delta_{u,v}(\{u, v\})) + w(\delta_{u,v}(\emptyset)) \tag{3.51}$$

- General non-negative weighted graph $G = (V, E, w)$, define $w(\delta(\cdot))$:

$$f(X) = w(\delta(X)) = \sum_{(u,v) \in E(G)} w(\delta_{u,v}(X \cap \{u, v\})) \tag{3.52}$$

- This is easily shown to be submodular using properties we will soon see (namely, submodularity closed under summation and restriction).

Other graph functions that are submodular/supermodular

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

- Let $V(X)$ be the vertices adjacent to some edge in $X \subseteq E(G)$, then $|V(X)|$ (the vertex function) is submodular.

Other graph functions that are submodular/supermodular

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

- Let $V(X)$ be the vertices adjacent to some edge in $X \subseteq E(G)$, then $|V(X)|$ (the vertex function) is submodular.
- Let $E(S)$ be the edges with both vertices in $S \subseteq V(G)$. Then $|E(S)|$ (the interior edge function) is supermodular.

Other graph functions that are submodular/supermodular

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

- Let $V(X)$ be the vertices adjacent to some edge in $X \subseteq E(G)$, then $|V(X)|$ (the vertex function) is submodular.
- Let $E(S)$ be the edges with both vertices in $S \subseteq V(G)$. Then $|E(S)|$ (the interior edge function) is supermodular.
- Let $I(S)$ be the edges with at least one vertex in $S \subseteq V(G)$. Then $|I(S)|$ (the incidence function) is submodular.

Other graph functions that are submodular/supermodular

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

- Let $V(X)$ be the vertices adjacent to some edge in $X \subseteq E(G)$, then $|V(X)|$ (the vertex function) is submodular.
- Let $E(S)$ be the edges with both vertices in $S \subseteq V(G)$. Then $|E(S)|$ (the interior edge function) is supermodular.
- Let $I(S)$ be the edges with at least one vertex in $S \subseteq V(G)$. Then $|I(S)|$ (the incidence function) is submodular.
- Recall $|\delta(S)|$, is the set size of edges with exactly one vertex in $S \subseteq V(G)$ is submodular (cut size function). Thus, we have $I(S) = E(S) \cup \delta(S)$ and $E(S) \cap \delta(S) = \emptyset$, and thus that $|I(S)| = |E(S)| + |\delta(S)|$.

Other graph functions that are submodular/supermodular

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

- Let $V(X)$ be the vertices adjacent to some edge in $X \subseteq E(G)$, then $|V(X)|$ (the vertex function) is submodular.
- Let $E(S)$ be the edges with both vertices in $S \subseteq V(G)$. Then $|E(S)|$ (the interior edge function) is supermodular.
- Let $I(S)$ be the edges with at least one vertex in $S \subseteq V(G)$. Then $|I(S)|$ (the incidence function) is submodular.
- Recall $|\delta(S)|$, is the set size of edges with exactly one vertex in $S \subseteq V(G)$ is submodular (cut size function). Thus, we have $I(S) = E(S) \cup \delta(S)$ and $E(S) \cap \delta(S) = \emptyset$, and thus that $|I(S)| = |E(S)| + |\delta(S)|$. So we can get a submodular function by summing a submodular and a supermodular function.

Other graph functions that are submodular/supermodular

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

- Let $V(X)$ be the vertices adjacent to some edge in $X \subseteq E(G)$, then $|V(X)|$ (the vertex function) is submodular.
- Let $E(S)$ be the edges with both vertices in $S \subseteq V(G)$. Then $|E(S)|$ (the interior edge function) is supermodular.
- Let $I(S)$ be the edges with at least one vertex in $S \subseteq V(G)$. Then $|I(S)|$ (the incidence function) is submodular.
- Recall $|\delta(S)|$, is the set size of edges with exactly one vertex in $S \subseteq V(G)$ is submodular (cut size function). Thus, we have $I(S) = E(S) \cup \delta(S)$ and $E(S) \cap \delta(S) = \emptyset$, and thus that $|I(S)| = |E(S)| + |\delta(S)|$. So we can get a submodular function by summing a submodular and a supermodular function. If you had to guess, is this always the case?

Other graph functions that are submodular/supermodular

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

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

Number of connected components in a graph via edges

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

Number of connected components in a graph via edges

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

Number of connected components in a graph via edges

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

Number of connected components in a graph via edges

- Recall, $f : 2^V \to \mathbb{R}$ is submodular, then so is $\bar{f} : 2^V \to \mathbb{R}$ defined as $\bar{f}(S) = f(V \setminus S)$.
- Hence, if $f : 2^V \to \mathbb{R}$ is supermodular, then so is $\bar{f} : 2^V \to \mathbb{R}$ defined as $\bar{f}(S) = f(V \setminus S)$.
- Given a graph $G = (V, E)$, for each $A \subseteq E(G)$, let $c(A)$ denote the number of connected components of the (spanning) subgraph $(V(G), A)$, with $c : 2^E \to \mathbb{R}_+$.
- $c(A)$ is monotone non-increasing, $c(A + a) - c(A) \leq 0$.

Number of connected components in a graph via edges

- Recall, $f : 2^V \to \mathbb{R}$ is submodular, then so is $\bar{f} : 2^V \to \mathbb{R}$ defined as $\bar{f}(S) = f(V \setminus S)$.
- Hence, if $f : 2^V \to \mathbb{R}$ is supermodular, then so is $\bar{f} : 2^V \to \mathbb{R}$ defined as $\bar{f}(S) = f(V \setminus S)$.
- Given a graph $G = (V, E)$, for each $A \subseteq E(G)$, let $c(A)$ denote the number of connected components of the (spanning) subgraph $(V(G), A)$, with $c : 2^E \to \mathbb{R}_+$.
- $c(A)$ is monotone non-increasing, $c(A + a) - c(A) \leq 0$.
- Then $c(A)$ is supermodular, i.e.,

$$c(A + a) - c(A) \leq c(B + a) - c(B) \tag{3.53}$$

with $A \subseteq B \subseteq E \setminus \{a\}$.

Number of connected components in a graph via edges

- Recall, $f : 2^V \to \mathbb{R}$ is submodular, then so is $\bar{f} : 2^V \to \mathbb{R}$ defined as $\bar{f}(S) = f(V \setminus S)$.
- Hence, if $f : 2^V \to \mathbb{R}$ is supermodular, then so is $\bar{f} : 2^V \to \mathbb{R}$ defined as $\bar{f}(S) = f(V \setminus S)$.
- Given a graph $G = (V, E)$, for each $A \subseteq E(G)$, let $c(A)$ denote the number of connected components of the (spanning) subgraph $(V(G), A)$, with $c : 2^E \to \mathbb{R}_+$.
- $c(A)$ is monotone non-increasing, $c(A + a) - c(A) \leq 0$.
- Then $c(A)$ is supermodular, i.e.,
$$c(A + a) - c(A) \leq c(B + a) - c(B) \tag{3.53}$$
with $A \subseteq B \subseteq E \setminus \{a\}$.
- Intuition: an edge is "more" (no less) able to bridge separate components (and reduce the number of conected components) when edge is added in a smaller context than when added in a larger context.

Number of connected components in a graph via edges

- Recall, $f : 2^V \to \mathbb{R}$ is submodular, then so is $\bar{f} : 2^V \to \mathbb{R}$ defined as $\bar{f}(S) = f(V \setminus S)$.
- Hence, if $f : 2^V \to \mathbb{R}$ is supermodular, then so is $\bar{f} : 2^V \to \mathbb{R}$ defined as $\bar{f}(S) = f(V \setminus S)$.
- Given a graph $G = (V, E)$, for each $A \subseteq E(G)$, let $c(A)$ denote the number of connected components of the (spanning) subgraph $(V(G), A)$, with $c : 2^E \to \mathbb{R}_+$.
- $c(A)$ is monotone non-increasing, $c(A + a) - c(A) \leq 0$.
- Then $c(A)$ is supermodular, i.e.,

$$c(A + a) - c(A) \leq c(B + a) - c(B) \tag{3.53}$$

  with $A \subseteq B \subseteq E \setminus \{a\}$.
- Intuition: an edge is "more" (no less) able to bridge separate components (and reduce the number of conected components) when edge is added in a smaller context than when added in a larger context.
- $\bar{c}(A) = c(E \setminus A)$ is the number of connected components in G when we remove A, so is also supermodular, but monotone non-decreasing.

Graph Strength

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

Graph Strength

- So $\bar{c}(A) = c(E \setminus A)$ is the number of connected components in G when we remove A, is supermodular.
- Maximizing $\bar{c}(A)$ might seem as a goal for a network attacker — many connected components means that many points in the network have lost connectivity to many other points (unprotected network).

Graph Strength

- So $\bar{c}(A) = c(E \setminus A)$ is the number of connected components in G when we remove A, is supermodular.
- Maximizing $\bar{c}(A)$ might seem as a goal for a network attacker — many connected components means that many points in the network have lost connectivity to many other points (unprotected network).
- If we can remove a small set A and shatter the graph into many connected components, then the graph is weak.

Graph Strength

- So $\bar{c}(A) = c(E \setminus A)$ is the number of connected components in G when we remove A, is supermodular.
- Maximizing $\bar{c}(A)$ might seem as a goal for a network attacker — many connected components means that many points in the network have lost connectivity to many other points (unprotected network).
- If we can remove a small set A and shatter the graph into many connected components, then the graph is weak.
- An attacker wishes to choose a small number of edges (since it is cheap) to shatter the graph into as many components as possible.

Graph Strength

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

Graph Strength

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

Graph Strength

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

$$w(A) = \sum_{e \in A} w_e \tag{3.54}$$

so that $w(E(G[S]))$ is the “internal strength” of the vertex set S.
Notation: S is a set of nodes, $G[S]$ is the vertex-induced subgraph of G induced by vertices S, $E(G[S])$ are the edges contained within this induced subgraph, and $w(E(G[S]))$ is the weight of these edges.

Graph Strength

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

$$w(A) = \sum_{e \in A} w_e \tag{3.54}$$

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 component" for a network attacker.

Graph Strength

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

$$w(A) = \sum_{e \in A} w_e \tag{3.54}$$

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 component" for a network attacker.
- A form of graph strength can then be defined as the following:

$$\textit{strength}(G, w) = \min_{A \subseteq E(G) : \bar{c}(A) > 1} \frac{w(A)}{\bar{c}(A) - 1} \tag{3.55}$$

Graph Strength

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

$$w(A) = \sum_{e \in A} w_e \tag{3.54}$$

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 component" for a network attacker.
- A form of graph strength can then be defined as the following:

$$\textit{strength}(G, w) = \min_{A \subseteq E(G): \bar{c}(A) > 1} \frac{w(A)}{\bar{c}(A) - 1} \tag{3.55}$$

- 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).

Graph Strength

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

$$w(A) = \sum_{e \in A} w_e \tag{3.54}$$

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 component" for a network attacker.
- A form of graph strength can then be defined as the following:

$$\textit{strength}(G, w) = \min_{A \subseteq E(G): \bar{c}(A) > 1} \frac{w(A)}{\bar{c}(A) - 1} \tag{3.55}$$

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

Submodularity, Quadratic Structures, and Cuts

Lemma 3.7.4

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*

$$f(X) = m^{\intercal}\mathbf{1}_X + \frac{1}{2}\mathbf{1}_X^{\intercal}\mathbf{M}\mathbf{1}_X \tag{3.56}$$

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.

Submodularity, Quadratic Structures, and Cuts

Lemma 3.7.4

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

$$f(X) = m^\intercal \mathbf{1}_X + \frac{1}{2}\mathbf{1}_X^\intercal \mathbf{M} \mathbf{1}_X \tag{3.56}$$

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

Proof.

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

Submodularity, Quadratic Structures, and Cuts

Lemma 3.7.4

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

$$f(X) = m^\intercal \mathbf{1}_X + \frac{1}{2}\mathbf{1}_X^\intercal \mathbf{M}\mathbf{1}_X \tag{3.56}$$

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 $-w(E(X))$ (non-positive modular) is submodular.
- f is a modular function $m^\intercal \mathbf{1}_A = m(A)$ added to a weighted submodular function, hence f is submodular.

Submodularity, Quadratic Structures, and Cuts

Proof of Lemma 3.7.4 cont.

- Conversely, suppose f is submodular.

□

Submodularity, Quadratic Structures, and Cuts

Proof of Lemma 3.7.4 cont.

- Conversely, suppose f is submodular.
- Then $f(\{u\}) + f(\{v\}) \geq f(\{u, v\}) + f(\emptyset)$ while $f(\emptyset) = 0$.

$\square$

Submodularity, Quadratic Structures, and Cuts

Proof of Lemma 3.7.4 cont.

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

$$0 \leq f(\{u\}) + f(\{v\}) - f(\{u,v\}) \tag{3.57}$$

$$= m(u) + \frac{1}{2}M_{u,u} + m(v) + \frac{1}{2}M_{v,v} \tag{3.58}$$

$$- \Big(m(u) + m(v) + \frac{1}{2}M_{u,u} + M_{u,v} + \frac{1}{2}M_{v,v}\Big) \tag{3.59}$$

$$= -M_{u,v} \tag{3.60}$$

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

$\square$

Set Cover and Maximum Coverage

- We are given a finite set V of n elements and a set of subsets $\mathcal{V} = \{V_1, V_2, \dots, V_m\}$ of m subsets of V, so that $V_i \subseteq V$ and $\bigcup_i V_i = V$.

Set Cover and Maximum Coverage

- We are given a finite set V of n elements and a set of subsets $\mathcal{V} = \{V_1, V_2, \dots, V_m\}$ of m subsets of V, so that $V_i \subseteq V$ and $\bigcup_i V_i = V$.
- The goal of minimum SET COVER is to choose the smallest subset $A \subseteq [m] \triangleq \{1, \dots, m\}$ such that $\bigcup_{a \in A} V_a = V$.

Set Cover and Maximum Coverage

- We are given a finite set V of n elements and a set of subsets $\mathcal{V} = \{V_1, V_2, \ldots, V_m\}$ of m subsets of V, so that $V_i \subseteq V$ and $\bigcup_i V_i = V$.
- The goal of minimum SET COVER is to choose the smallest subset $A \subseteq [m] \triangleq \{1, \ldots, m\}$ such that $\bigcup_{a \in A} V_a = V$.
- Maximum k cover: The goal in MAXIMUM COVERAGE is, given an integer $k \leq m$, select k subsets, say $\{a_1, a_2, \ldots, a_k\}$ with $a_i \in [m]$ such that $|\bigcup_{i=1}^k V_{a_i}|$ is maximized.

Set Cover and Maximum Coverage

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

Other Covers

Definition 3.7.5 (vertex cover)

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

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

Definition 3.7.6 (edge cover)

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

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

Graph Cut Problems

- 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

- MINIMUM CUT: Given a graph $G = (V, E)$, find a set of vertices $S \subseteq V$ that minimize the cut (set of edges) between S and $V \setminus S$.
- MAXIMUM CUT: Given a graph $G = (V, E)$, find a set of vertices $S \subseteq V$ that minimize the cut (set of edges) between S and $V \setminus S$.

Graph Cut Problems

- Minimum cut: Given a graph $G = (V, E)$, find a set of vertices $S \subseteq V$ that minimize the cut (set of edges) between S and $V \setminus S$.
- Maximum cut: Given a graph $G = (V, E)$, find a set of vertices $S \subseteq V$ that minimize the cut (set of edges) between S and $V \setminus S$.
- Let $f : 2^V \to \mathbb{R}_+$ be the cut function, namely for any given set of nodes $X \subseteq V$, $f(X)$ measures the number of edges between nodes X and $V \setminus X$.

Graph Cut Problems

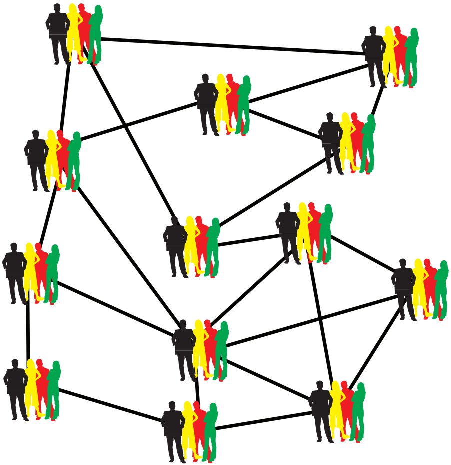
- MINIMUM CUT: Given a graph $G = (V, E)$, find a set of vertices $S \subseteq V$ that minimize the cut (set of edges) between S and $V \setminus S$.
- MAXIMUM CUT: Given a graph $G = (V, E)$, find a set of vertices $S \subseteq V$ that minimize the cut (set of edges) between S and $V \setminus S$.
-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

Graph Cut Problems

- MINIMUM CUT: Given a graph $G = (V, E)$, find a set of vertices $S \subseteq V$ that minimize the cut (set of edges) between S and $V \setminus S$.
- MAXIMUM CUT: Given a graph $G = (V, E)$, find a set of vertices $S \subseteq V$ that minimize the cut (set of edges) between S and $V \setminus S$.
- Let $f : 2^V \to \mathbb{R}_+$ be the cut function, namely for any given set of nodes $X \subseteq V$, $f(X)$ measures the number of edges between nodes X and $V \setminus X$.
- Weighted versions, where rather than count, we sum the (non-negative) weights of the edges of a cut.
- Many examples of this, we will see more later.

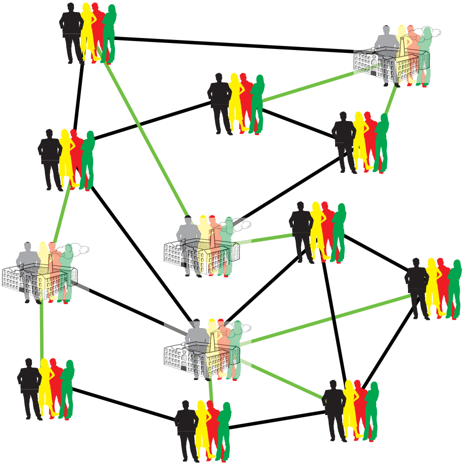
Facility/Plant Location (uncapacitated)

- 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.

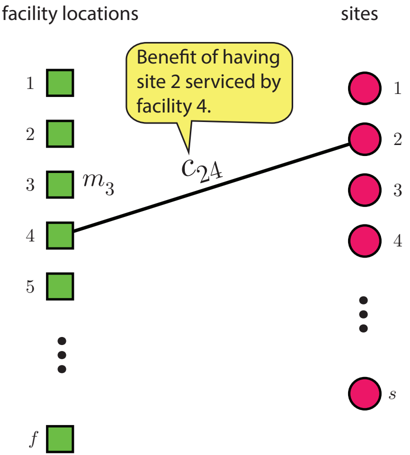


Facility/Plant Location (uncapacitated)

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

$$f(A) = \sum_{i \in F} \max_{j \in A} c_{ij}. \qquad (3.61)$$



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

Facility/Plant Location (uncapacitated) w. plant benefits

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

Facility/Plant Location (uncapacitated) w. plant benefits

- Let $F = \{1, \ldots, f\}$ be a set of possible factory/plant locations for facilities to be built.
- $S = \{1, \ldots, s\}$ is a set of sites (e.g., cities, clients) needing service.
- Let c_{ij} be the "benefit" (e.g., $1/c_{ij}$ is the cost) of servicing site i with facility location j.
- Let m_j be the benefit (e.g., either $1/m_j$ is the cost or $-m_j$ is the cost) to build a plant at location j.
- Each site should be serviced by only one plant but no less than one.

Facility/Plant Location (uncapacitated) w. plant benefits

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

Facility/Plant Location (uncapacitated) w. plant benefits

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

$$f(A) = \sum_{j \in A} m_j + \sum_{i \in F} \max_{j \in A} c_{ij}. \tag{3.62}$$

Facility/Plant Location (uncapacitated) w. plant benefits

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

$$f(A) = \sum_{j \in A} m_j + \sum_{i \in F} \max_{j \in A} c_{ij}. \tag{3.62}$$

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

Matrix Rank functions

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

Matrix Rank functions

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

$$x_v = \sum_{i=1}^{k} \alpha_i x_{v_i} \tag{3.63}$$

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

Matrix Rank functions

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

$$x_v = \sum_{i=1}^{k} \alpha_i x_{v_i} \tag{3.63}$$

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

Example: Rank function of a matrix

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

Skip matrix rank example

Example: Rank function of a matrix

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

Skip matrix rank example

Example: Rank function of a matrix

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

Skip matrix rank example

Example: Rank function of a matrix

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

Skip matrix rank example

Example: Rank function of a matrix

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

Skip matrix rank example

Example: Rank function of a matrix

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Rank function of a matrix

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

Rank function of a matrix

- Let $A, B \subseteq V$ be two subsets of column indices.
- The rank of the two sets unioned together $A \cup B$ is no more than the sum of the two individual ranks.

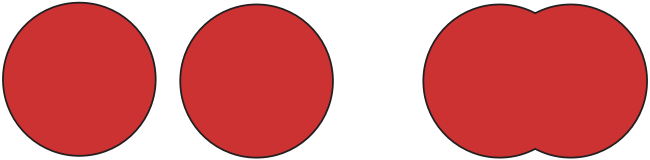
Rank function of a matrix

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

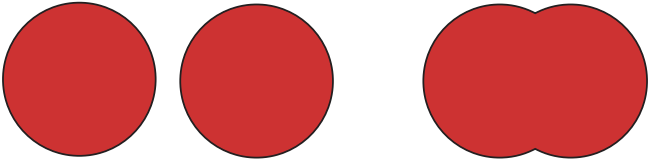
$$r(A) \quad + \quad r(B) \qquad \geq \qquad r(A \cup B)$$



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

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

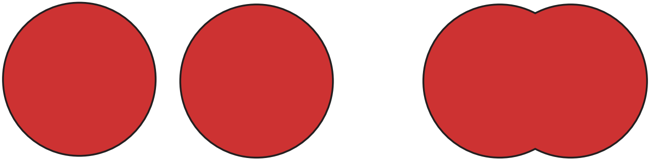


- If some of the dimensions spanned by A overlap some of the dimensions spanned by B (i.e., if $\exists$ common span), then that area is counted twice in $r(A) + r(B)$, so the inequality will be strict.

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

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



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

Rank functions of a matrix

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

Rank functions of a matrix

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

Rank functions of a matrix

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

Rank functions of a matrix

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

Rank functions of a matrix

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

Rank functions of a matrix

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

Rank functions of a matrix

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

$$r(A) + r(B) = r(A_r) + 2r(C) + r(B_r). \tag{3.64}$$

Rank functions of a matrix

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

$$r(A) + r(B) = r(A_r) + 2r(C) + r(B_r). \tag{3.64}$$

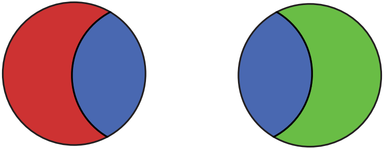
- But $r(A \cup B)$ counts the dimensions spanned by C only once.

$$r(A \cup B) = r(A_r) + r(C) + r(B_r) \tag{3.65}$$

Rank functions of a matrix

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

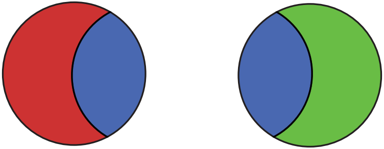
$$r(A) + r(B) = r(A_r) + 2r(C) + r(B_r)$$



Rank functions of a matrix

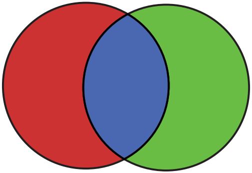
- Then $r(A) + r(B)$ counts the dimensions spanned by C twice, i.e.,

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



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

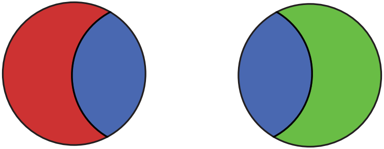
$$r(A \cup B) = r(A_r) + r(C) + r(B_r)$$



Rank functions of a matrix

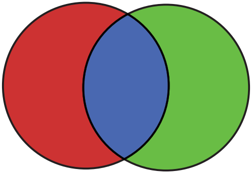
- Then $r(A) + r(B)$ counts the dimensions spanned by C twice, i.e.,

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



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

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

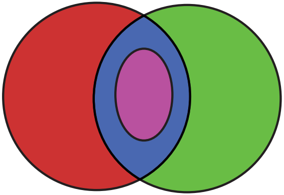


- Thus, we have subadditivity: $r(A) + r(B) \geq r(A \cup B)$. Can we add more to the r.h.s. and still have an inequality? Yes.

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).

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

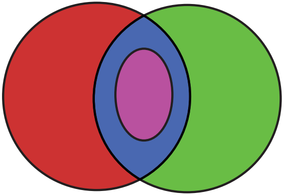


In short:

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).

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



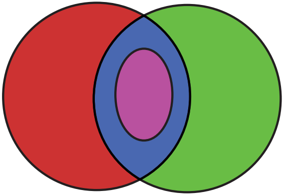
In short:

- 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).

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

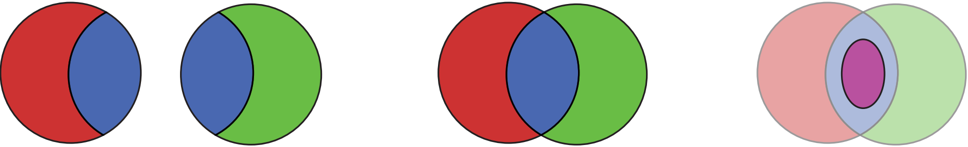


In short:

- 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 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.

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,

$$f(X) = r(\cup_{s \in S} X_s) \tag{3.66}$$

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,

$$f(X) = r(\cup_{s \in S} X_s) \tag{3.66}$$

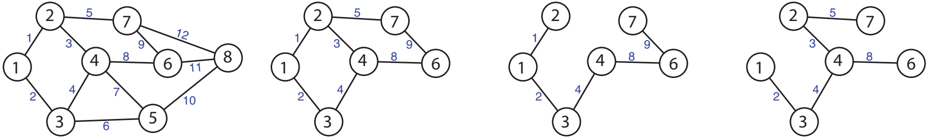
  we have that f is submodular, and is known to be a polymatroid rank function.
- In general (as we will see) polymatroid rank functions are submodular, normalized $f(\emptyset) = 0$, and monotone non-decreasing ($f(A) \leq f(B)$ whenever $A \subseteq B$).

Spanning trees

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

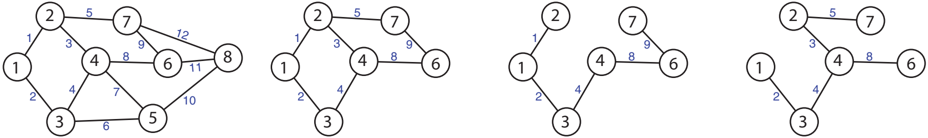
Spanning trees

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



Spanning trees

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



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