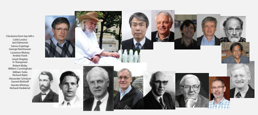
Submodular Functions, Optimization, and Applications to Machine Learning
— Spring Quarter, Lecture 5 —

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

Prof. Jeff Bilmes

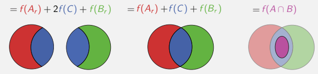
University of Washington, Seattle
Department of Electrical Engineering
http://melodi.ee.washington.edu/~bilmes

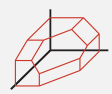
April 9th, 2018



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

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





Cumulative Outstanding Reading

- Read chapter 1 from Fujishige's book.
- Read chapter 2 from Fujishige's book.

Announcements, Assignments, and Reminders

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

Class Road Map - EE563

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

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

Summary submodular properties

- $c(A)$, number of connected components induced by $A \subseteq E(G)$ is supermodular.
- $f(X) = m^\intercal \mathbf{1}_X + \frac{1}{2}\mathbf{1}_X^\intercal \mathbf{M}\mathbf{1}_X$ submodular iff off-diagonal elements of M non-positive.
- Weighted set cover $f(A) = w(\bigcup_{a \in A} U_a)$, other cover functions, cut functions.
- Matrix rank $r(A)$, the dimensionality of the vector space spanned by the set of vectors $\{x_a\}_{a \in A}$.
- Adding modular functions to submodular functions preserves submodularity.
- Conic mixtures: if $\alpha_i \geq 0$ and $f_i : 2^V \to \mathbb{R}$ is submodular, then so is $\sum_i \alpha_i f_i$.
- Restrictions: $f'(A) = f(A \cap S)$
- max: $f(A) = \max_{j \in A} c_j$ and facility location.
- Log determinant $f(A) = \log\det(\mathbf{\Sigma}_A)$

Concave over non-negative modular

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

$$f(A) = \phi(m(A)) \tag{5.1}$$

then f is submodular.

Proof.

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

$$\phi(m(A) + m(v)) - \phi(m(A)) \geq \phi(m(B) + m(v)) - \phi(m(B)) \tag{5.2}$$

□

A form of converse is true as well.

Concave composed with non-negative modular

Theorem 5.3.1

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

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

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

$$f(A) = \sum_{i=1}^{K} \phi_i(m_i(A)) \tag{5.3}$$

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

Monotonicity

Definition 5.3.2

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

Definition 5.3.3

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

Composition of non-decreasing submodular and non-decreasing concave

Theorem 5.3.4

Given two functions, one defined on sets

$$f : 2^V \to \mathbb{R} \tag{5.4}$$

and another continuous valued one:

$$\phi : \mathbb{R} \to \mathbb{R} \tag{5.5}$$

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

Monotone difference of two functions

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

$$h(A) = \min(f(A), g(A)) \tag{5.6}$$

is submodular.

Proof.

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

$$h(X) + h(Y) = f(X) + f(Y) \geq f(X \cup Y) + f(X \cap Y) \tag{5.7}$$

or

$$h(X) + h(Y) = g(X) + g(Y) \geq g(X \cup Y) + g(X \cap Y), \tag{5.8}$$

the result (Equation 5.6 being submodular) follows since

$$\begin{matrix} f(X) + f(Y) \\ g(X) + g(Y) \end{matrix} \geq \min(f(X \cup Y), g(X \cup Y)) + \min(f(X \cap Y), g(X \cap Y)) \tag{5.9}$$

...

Monotone difference of two functions

...cont.

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

$$h(X) + h(Y) = f(X) + g(Y) \geq f(X \cup Y) + f(X \cap Y) + g(Y) - f(Y) \tag{5.10}$$

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

$$h(X) + h(Y) \geq g(X \cup Y) + f(X \cap Y) \geq h(X \cup Y) + h(X \cap Y) \tag{5.11}$$

□

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

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

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

$$h(A) = \min(\alpha, f(A)) \tag{5.12}$$

is submodular.

Proof.

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

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

More on Min - the saturate trick

- In general, the minimum of two submodular functions is not submodular (unlike concave functions, closed under min).
- However, when wishing to maximize two monotone non-decreasing submodular functions f, g, we can define function $h_\alpha : 2^V \to \mathbb{R}$ as

$$h_\alpha(A) = \frac{1}{2}\Big(\min(\alpha, f(A)) + \min(\alpha, g(A))\Big) \tag{5.13}$$

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

Arbitrary functions: difference between submodular funcs.

Theorem 5.3.5

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

Proof.

Let h be given and arbitrary, and define:

$$\alpha \triangleq \min_{X,Y : X \not\subseteq Y, Y \not\subseteq X} \Big(h(X) + h(Y) - h(X \cup Y) - h(X \cap Y)\Big) \tag{5.14}$$

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

$$\beta \triangleq \min_{X,Y : X \not\subseteq Y, Y \not\subseteq X} \Big(f(X) + f(Y) - f(X \cup Y) - f(X \cap Y)\Big). \tag{5.15}$$

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

Arbitrary functions as difference between submodular funcs.

...cont.

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

$$h'(A) = h(A) + \frac{|\alpha|}{\beta} f(A) \tag{5.16}$$

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

□

Gain

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

$$\begin{aligned} f(A \cup \{j\}) - f(A) &\triangleq \rho_j(A) && (5.17) \\ &\triangleq \rho_A(j) && (5.18) \\ &\triangleq \nabla_j f(A) && (5.19) \\ &\triangleq f(\{j\}|A) && (5.20) \\ &\triangleq f(j|A) && (5.21) \end{aligned}$$

- We'll use $f(j|A)$.
- Submodularity's diminishing returns definition can be stated as saying that $f(j|A)$ is a monotone non-increasing function of A, since $f(j|A) \geq f(j|B)$ whenever $A \subseteq B$ (conditioning reduces valuation).

Gain Notation

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

$$f(A|B) \triangleq f(A \cup B) - f(B) \tag{5.22}$$

So when j is any singleton

$$f(j|B) = f(\{j\}|B) = f(\{j\} \cup B) - f(B) \tag{5.23}$$

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

Totally normalized functions

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

$$\bar{g}(A) = g(A) - \sum_{a \in A} g(a|V \setminus \{a\}) = g(A) - m_g(A) \tag{5.24}$$

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

$$\bar{g}(v|A) = g(v|A) - g(v|V \setminus \{v\}) \geq 0 \tag{5.25}$$

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

Arbitrary function as difference between two polymatroids

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

$$\begin{aligned} h = f - g &= \bar{f} + m_f - (\bar{g} + m_g) && (5.26)\\ &= \bar{f} - \bar{g} + (m_f - m_g) && (5.27)\\ &= \bar{f} - \bar{g} + m_{f-h} && (5.28)\\ &= \bar{f} + m^+_{f-g} - (\bar{g} + (-m_{f-g})^+) && (5.29)\end{aligned}$$

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

Two Equivalent Submodular Definitions

Definition 5.4.1 (submodular concave)

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

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B) \quad (5.8)$$

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

Definition 5.4.2 (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:

$$f(A \cup \{v\}) - f(A) \geq f(B \cup \{v\}) - f(B) \quad (5.9)$$

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

Submodular Definition: Group Diminishing Returns

An alternate and equivalent definition is:

Definition 5.4.1 (group diminishing returns)

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

$$f(A \cup C) - f(A) \geq f(B \cup C) - f(B) \tag{5.30}$$

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

Submodular Definition Basic Equivalencies

We want to show that Submodular Concave (Definition 5.4.1), Diminishing Returns (Definition 5.4.2), and Group Diminishing Returns (Definition 5.4.1) are identical. We will show that:

- Submodular Concave $\Rightarrow$ Diminishing Returns
- Diminishing Returns $\Rightarrow$ Group Diminishing Returns
- Group Diminishing Returns $\Rightarrow$ Submodular Concave

Submodular Concave $\Rightarrow$ Diminishing Returns

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

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

$$f(A + v) + f(B) \geq f(B + v) + f(A) \tag{5.31}$$

- Rearranging, we have

$$f(A + v) - f(A) \geq f(B + v) - f(B) \tag{5.32}$$

□

Diminishing Returns $\Rightarrow$ Group Diminishing Returns

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

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

$$f(A \cup C) - f(A) \tag{5.33}$$

$$= f(A \cup C) - \sum_{i=1}^{k-1} \Big(f(A \cup \{c_1, \ldots, c_i\}) - f(A \cup \{c_1, \ldots, c_i\}) \Big) - f(A) \tag{5.34}$$

$$= \sum_{i=1}^{k} \Big(f(A \cup \{c_1 \ldots c_i\}) - f(A \cup \{c_1 \ldots c_{i-1}\}) \Big) = \sum_{i=1}^{k} f(c_i | A \cup \{c_1 \ldots c_{i-1}\}) \tag{5.35}$$

$$\geq \sum_{i=1}^{k} f(c_i | B \cup \{c_1 \ldots c_{i-1}\}) = \sum_{i=1}^{k} \Big(f(B \cup \{c_1 \ldots c_i\}) - f(B \cup \{c_1 \ldots c_{i-1}\}) \Big) \tag{5.36}$$

$$= f(B \cup C) - \sum_{i=1}^{k-1} \Big(f(B \cup \{c_1, \ldots, c_i\}) - f(B \cup \{c_1, \ldots, c_i\}) \Big) - f(B) \tag{5.37}$$

$$= f(B \cup C) - f(B) \tag{5.38}$$

□

Group Diminishing Returns $\Rightarrow$ Submodular Concave

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

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

$$f(A'+C) - f(A') \geq f(B'+C) - f(B') \tag{5.39}$$

giving

$$f(A'+C) + f(B') \geq f(B'+C) + f(A') \tag{5.40}$$

or

$$f(A \cap B + A \setminus B) + f(B) \geq f(B + A \setminus B) + f(A \cap B) \tag{5.41}$$

which is the same as the submodular concave condition

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B) \tag{5.42}$$

Submodular Definition: Four Points

Definition 5.4.2 ("singleton", or "four points")

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

$$f(A \cup \{a\}) + f(A \cup \{b\}) \geq f(A \cup \{a,b\}) + f(A) \tag{5.43}$$

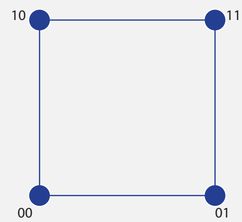
This follows immediately from diminishing returns. To achieve diminishing returns, assume $A \subset B$ with $B \setminus A = \{b_1, b_2, \ldots, b_k\}$. Then

$$\begin{aligned} f(A+a) - f(A) &\geq f(A+b_1+a) - f(A+b_1) && (5.44)\\ &\geq f(A+b_1+b_2+a) - f(A+b_1+b_2) && (5.45)\\ &\geq \ldots && (5.46)\\ &\geq f(A+b_1+\cdots+b_k+a) - f(A+b_1+\cdots+b_k) && (5.47)\\ &= f(B+a) - f(B) && (5.48)\end{aligned}$$

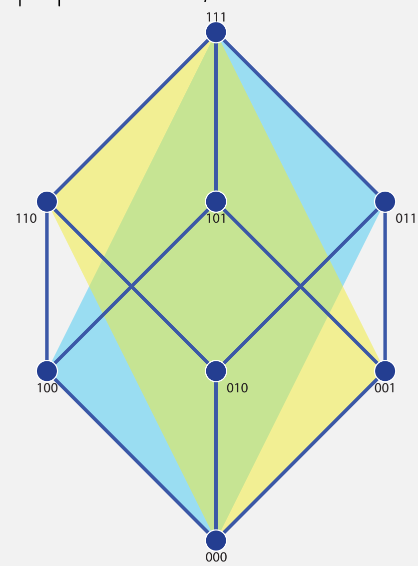
Submodular on Hypercube Vertices

- Test submodularity via values on verticies of hypercube.

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



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



How many inequalities?

Submodular Concave ≡ Diminishing Returns, in one slide.

Theorem 5.4.3

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

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

if and only if

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

Proof.

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

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

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

Applying telescoping summation to both sides, we get:

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

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

□

Submodular bounds of a difference of comparable sets

- Given submodular f, and given you have $C, D \subseteq V$ with either $D \supseteq C$ or $D \subseteq C$ (comparable sets), and have an expression of the form:

$$f(C) - f(D) \tag{5.49}$$

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

$$f(C) - f(D) = f(C) - f(C \cup X) \geq f(C \cap X) - f(X) \tag{5.50}$$

or

$$f(C \cup X | C) \leq f(X | C \cap X) \tag{5.51}$$

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

$$f(C) - f(D) = f(C) - f(C \cap Y) \geq f(C \cup Y) - f(Y) \tag{5.52}$$

or

$$f(C | C \cap Y) \geq f(C \cup Y | Y) \tag{5.53}$$

- Equations (5.51) and (5.53) have same form.

Many (Equivalent) Definitions of Submodularity

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B), \quad \forall A, B \subseteq V \tag{5.54}$$

$$f(j|S) \geq f(j|T), \ \forall S \subseteq T \subseteq V, \text{ with } j \in V \setminus T \tag{5.55}$$

$$f(C|S) \geq f(C|T), \forall S \subseteq T \subseteq V, \text{ with } C \subseteq V \setminus T \tag{5.56}$$

$$f(j|S) \geq f(j|S \cup \{k\}), \ \forall S \subseteq V \text{ with } j \in V \setminus (S \cup \{k\}) \tag{5.57}$$

$$f(A \cup B | A \cap B) \leq f(A | A \cap B) + f(B | A \cap B), \quad \forall A, B \subseteq V \tag{5.58}$$

$$f(T) \leq f(S) + \sum_{j \in T \setminus S} f(j|S) - \sum_{j \in S \setminus T} f(j | S \cup T - \{j\}), \ \forall S, T \subseteq V \tag{5.59}$$

$$f(T) \leq f(S) + \sum_{j \in T \setminus S} f(j|S), \ \forall S \subseteq T \subseteq V \tag{5.60}$$

$$f(T) \leq f(S) - \sum_{j \in S \setminus T} f(j | S \setminus \{j\}) + \sum_{j \in T \setminus S} f(j | S \cap T) \ \forall S, T \subseteq V \tag{5.61}$$

$$f(T) \leq f(S) - \sum_{j \in S \setminus T} f(j | S \setminus \{j\}), \ \forall T \subseteq S \subseteq V \tag{5.62}$$

Equivalent Definitions of Submodularity

We've already seen that Eq. 5.54 $\equiv$ Eq. 5.55 $\equiv$ Eq. 5.56 $\equiv$ Eq. 5.57 $\equiv$ Eq. 5.58.

We next show that Eq. 5.57 $\Rightarrow$ Eq. 5.59 $\Rightarrow$ Eq. 5.60 $\Rightarrow$ Eq. 5.57.

Approach

To show these next results, we essentially first use:

$$f(S \cup T) = f(S) + f(T|S) \leq f(S) + \text{upper-bound} \tag{5.63}$$

and

$$f(T) + \text{lower-bound} \leq f(T) + f(S|T) = f(S \cup T) \tag{5.64}$$

leading to

$$f(T) + \text{lower-bound} \leq f(S) + \text{upper-bound} \tag{5.65}$$

or

$$f(T) \leq f(S) + \text{upper-bound} - \text{lower-bound} \tag{5.66}$$

Eq. 5.57 $\Rightarrow$ Eq. 5.59

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

$$f(S \cup T) - f(S) = \sum_{t=1}^{r} \Big(f(S \cup \{j_1, \ldots, j_t\}) - f(S \cup \{j_1, \ldots, j_{t-1}\}) \Big) \tag{5.67}$$

$$= \sum_{t=1}^{r} f(j_t | S \cup \{j_1, \ldots, j_{t-1}\}) \le \sum_{t=1}^{r} f(j_t | S) \tag{5.68}$$

$$= \sum_{j \in T \setminus S} f(j|S) \tag{5.69}$$

or

$$f(T|S) \le \sum_{j \in T \setminus S} f(j|S) \tag{5.70}$$

Examples and Properties Other Submodular Defs. Independence Matroids

Eq. 5.57 $\Rightarrow$ Eq. 5.59

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

$$f(S \cup T) - f(T) = \sum_{t=1}^{q} [f(T \cup \{k_1, \ldots, k_t\}) - f(T \cup \{k_1, \ldots, k_{t-1}\})] \tag{5.71}$$

$$= \sum_{t=1}^{q} f(k_t | T \cup \{k_1, \ldots, k_t\} \setminus \{k_t\}) \ge \sum_{t=1}^{q} f(k_t | T \cup S \setminus \{k_t\}) \tag{5.72}$$

$$= \sum_{j \in S \setminus T} f(j | S \cup T \setminus \{j\}) \tag{5.73}$$

Eq. 5.57 $\Rightarrow$ Eq. 5.59

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

$$f(T|S) = f(S \cup T) - f(S) \leq \sum_{j \in T \setminus S} f(j|S) \tag{5.74}$$

and the lower bound

$$f(S|T) = f(S \cup T) - f(T) \geq \sum_{j \in S \setminus T} f(j|S \cup T \setminus \{j\}) \tag{5.75}$$

This gives upper and lower bounds of the form

$$f(T) + \text{lower bound} \leq f(S \cup T) \leq f(S) + \text{upper bound}, \tag{5.76}$$

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

Eq. 5.59 $\Rightarrow$ Eq. 5.60

This follows immediately since if $S \subseteq T$, then $S \setminus T = \emptyset$, and the last term of Eq. 5.59 vanishes.

Many (Equivalent) Definitions of Submodularity

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B), \quad \forall A, B \subseteq V \tag{5.54}$$

$$f(j|S) \geq f(j|T), \; \forall S \subseteq T \subseteq V, \text{ with } j \in V \setminus T \tag{5.55}$$

$$f(C|S) \geq f(C|T), \forall S \subseteq T \subseteq V, \text{ with } C \subseteq V \setminus T \tag{5.56}$$

$$f(j|S) \geq f(j|S \cup \{k\}), \; \forall S \subseteq V \text{ with } j \in V \setminus (S \cup \{k\}) \tag{5.57}$$

$$f(A \cup B|A \cap B) \leq f(A|A \cap B) + f(B|A \cap B), \quad \forall A, B \subseteq V \tag{5.58}$$

$$f(T) \leq f(S) + \sum_{j \in T \setminus S} f(j|S) - \sum_{j \in S \setminus T} f(j|S \cup T - \{j\}), \; \forall S, T \subseteq V \tag{5.59}$$

$$f(T) \leq f(S) + \sum_{j \in T \setminus S} f(j|S), \; \forall S \subseteq T \subseteq V \tag{5.60}$$

$$f(T) \leq f(S) - \sum_{j \in S \setminus T} f(j|S \setminus \{j\}) + \sum_{j \in T \setminus S} f(j|S \cap T) \; \forall S, T \subseteq V \tag{5.61}$$

$$f(T) \leq f(S) - \sum_{j \in S \setminus T} f(j|S \setminus \{j\}), \; \forall T \subseteq S \subseteq V \tag{5.62}$$

Eq. 5.60 $\Rightarrow$ Eq. 5.57

Here, we set $T = S \cup \{j, k\}$, $j \notin S \cup \{k\}$ into Eq. 5.60 to obtain

$$\begin{aligned}
f(S \cup \{j, k\}) &\leq f(S) + f(j|S) + f(k|S) && (5.77)\\
&= f(S) + f(S + \{j\}) - f(S) + f(S + \{k\}) - f(S) && (5.78)\\
&= f(S + \{j\}) + f(S + \{k\}) - f(S) && (5.79)\\
&= f(j|S) + f(S + \{k\}) && (5.80)
\end{aligned}$$

giving

$$\begin{aligned}
f(j|S \cup \{k\}) &= f(S \cup \{j, k\}) - f(S \cup \{k\}) && (5.81)\\
&\leq f(j|S) && (5.82)
\end{aligned}$$

Submodular Concave

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

$$(\nabla_B f)(A) \triangleq f(A \cup B) - f(A \setminus B) = f\big(B|(A \setminus B)\big) \tag{5.83}$$

  read as: the derivative of f at A in the direction B.
- Hence, if $A \cap B = \emptyset$, then $(\nabla_B f)(A) = f(B|A)$.
- Consider a form of second derivative or 2nd difference:

$$\begin{aligned}(\nabla_C \nabla_B f)(A) &= \nabla_C[\, \overbrace{f(A \cup B) - f(A \setminus B)}^{(\nabla_B f)(A)}\,] &(5.84)\\ &= (\nabla_B f)(A \cup C) - (\nabla_B f)(A \setminus C) &(5.85)\\ &= f(A \cup B \cup C) - f((A \cup C) \setminus B)\\ &\quad - f((A \setminus C) \cup B) + f((A \setminus C) \setminus B) &(5.86)\end{aligned}$$

Submodular Concave

- If the second difference operator everywhere nonpositive:

$$\begin{aligned}&f(A \cup B \cup C) - f((A \cup C) \setminus B)\\ &\qquad - f((A \setminus C) \cup B) + f(A \setminus C \setminus B) \leq 0 \end{aligned} \tag{5.87}$$

  then we have the equation:

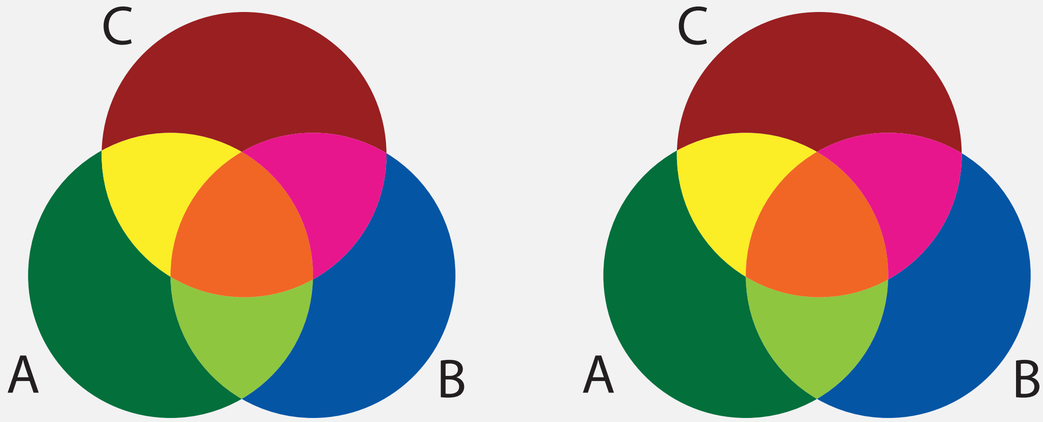
$$f((A \cup C) \setminus B) + f((A \setminus C) \cup B) \geq f(A \cup B \cup C) + f(A \setminus C \setminus B) \tag{5.88}$$

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

$$f(A') + f(B') \geq f(A' \cup B') + f(A' \cap B') \tag{5.89}$$

  and note that A' and B' so defined can be arbitrary.
- One sense in which submodular functions are like concave functions.

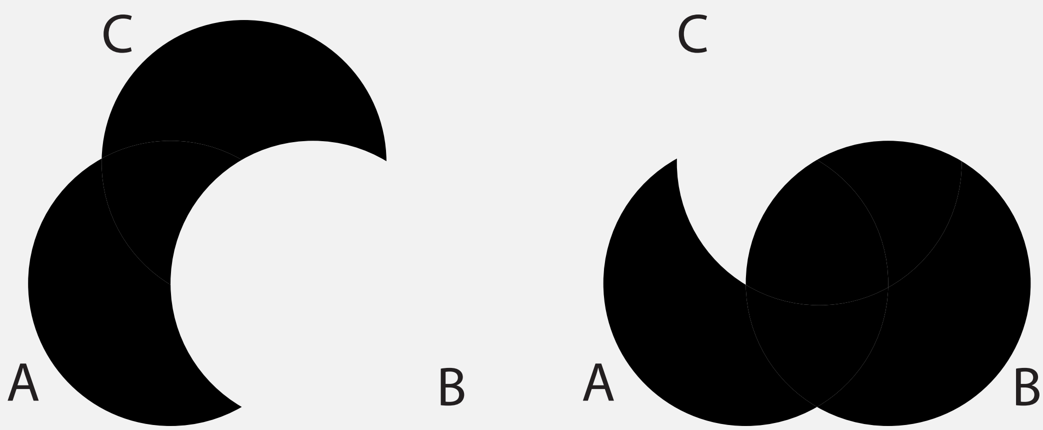
Submodular Concave



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

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

Submodular Concave



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

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

Submodularity and Concave

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

$$f(X + j) + f(X + k) \geq f(X + j + k) + f(X) \tag{5.90}$$

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

$$\nabla_j \nabla_k f(X) \leq 0 \tag{5.91}$$

Example: Rank function of a matrix

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

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

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

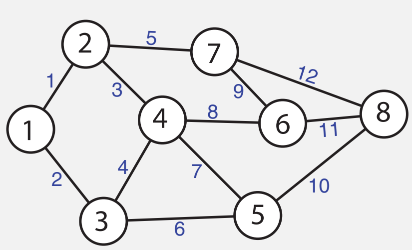
On Rank

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

Spanning trees/forests

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

$$x_{v,e} = \begin{cases} 1 & \text{if } v \in e \\ 0 & \text{if } v \notin e \end{cases} \tag{5.92}$$



$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \end{array} \begin{array}{c} \begin{array}{cccccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \end{array} \\ \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix} \end{array} \tag{5.93}$$

Spanning trees/forests & incidence matrices

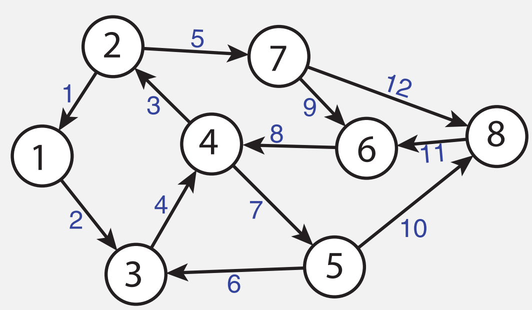
- We are given a graph $G = (V, E)$, we can arbitrarily orient the graph (make it directed) consider again the edges $E = E(G)$ as an index set.
- Consider instead the $|V| \times |E|$ incidence matrix of undirected graph G, which is the matrix $\mathbf{X}_G = (x_{v,e})_{v \in V(G), e \in E(G)}$ where

$$x_{v,e} = \begin{cases} 1 & \text{if } v \in e^+ \\ -1 & \text{if } v \in e^- \\ 0 & \text{if } v \notin e \end{cases} \tag{5.94}$$

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

Spanning trees/forests & incidence matrices

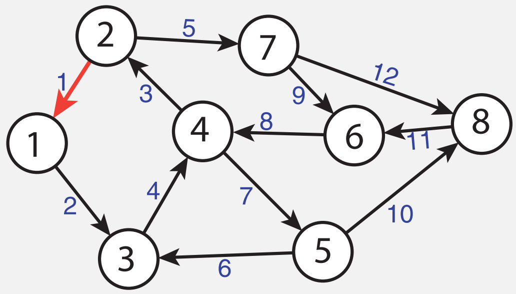
- A directed version of the graph (right) and its adjacency matrix (below).
- Orientation can be arbitrary.
- Note, rank of this matrix is 7.



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

Spanning trees

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



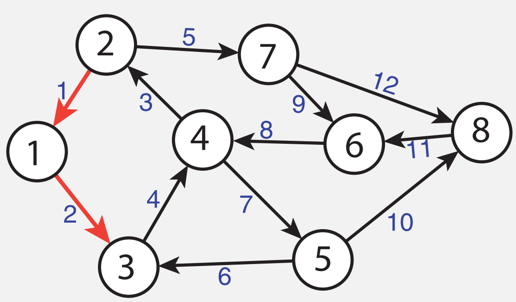
$$\begin{array}{c|c} & 1 \\ \begin{matrix}1\\2\\3\\4\\5\\6\\7\\8\end{matrix} & \begin{pmatrix}-1\\1\\0\\0\\0\\0\\0\\0\end{pmatrix}\end{array} \tag{5.95}$$

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

Examples and Properties | Other Submodular Defs. | Independence | Matroids

Spanning trees

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

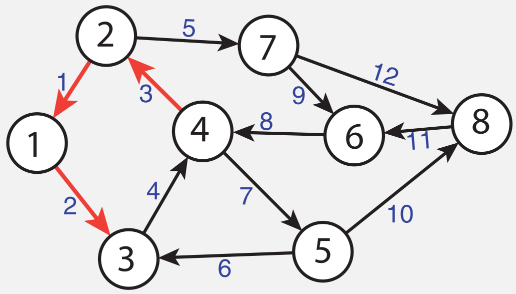


$$\begin{array}{c|cc} & 1 & 2 \\ \begin{matrix}1\\2\\3\\4\\5\\6\\7\\8\end{matrix} & \begin{pmatrix}-1 & 1\\1 & 0\\0 & -1\\0 & 0\\0 & 0\\0 & 0\\0 & 0\\0 & 0\end{pmatrix}\end{array} \tag{5.95}$$

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

Spanning trees

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

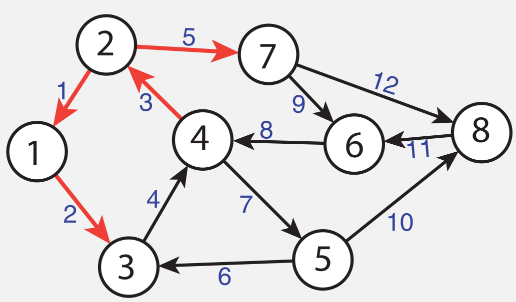


$$\begin{array}{c} \\ 1\\2\\3\\4\\5\\6\\7\\8 \end{array}\begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{pmatrix} -1 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \end{array} \tag{5.95}$$

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

Spanning trees

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

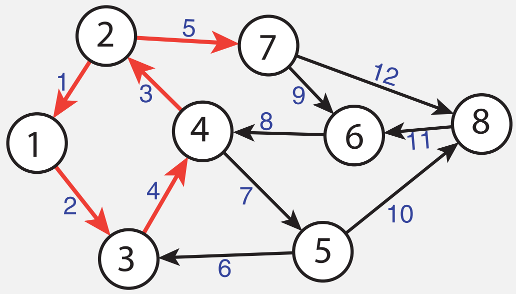


$$\begin{array}{c} \\ 1\\2\\3\\4\\5\\6\\7\\8 \end{array}\begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 5 \end{array} \\ \begin{pmatrix} -1 & 1 & 0 & 0 \\ 1 & 0 & -1 & 1 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \end{array} \tag{5.95}$$

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

Spanning trees

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

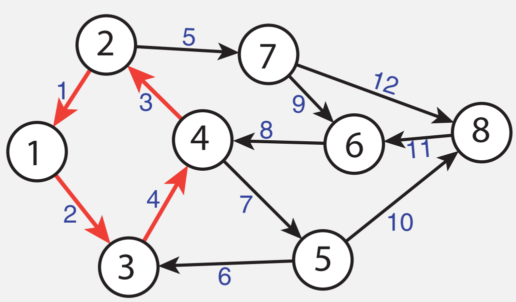


$$\begin{array}{c} \\ 1\\2\\3\\4\\5\\6\\7\\8 \end{array}\begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} -1 & 1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 1 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array} \tag{5.95}$$

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

Spanning trees

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



$$\begin{array}{c} \\ 1\\2\\3\\4\\5\\6\\7\\8 \end{array}\begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} -1 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \end{array} \tag{5.95}$$

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

Spanning trees, rank, and connected components

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

Spanning Tree Algorithms

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

Algorithm 2: Kruskal's Algorithm

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

Spanning Tree Algorithms

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

Algorithm 3: Jarník/Prim/Dijkstra Algorithm

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

Examples and Properties · Other Submodular Defs. · Independence · Matroids

Spanning Tree Algorithms

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

Algorithm 4: Borůvka's Algorithm

$F \leftarrow \emptyset$ /* We build up the edges of a forest in F */
while $G(V, F)$ *is disconnected* **do**
    forall *components* C_i *of* F **do**
        $F \leftarrow F \cup \{e_i\}$ for $e_i =$ the min-weight edge out of C_i;

Spanning Tree Algorithms

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

From Matrix Rank → Matroid

- So V is set of column vector indices of a matrix.
- Let $\mathcal{I} = \{I_1, I_2, \dots\}$ be a set of all subsets of V such that for any $I \in \mathcal{I}$, the vectors indexed by I are linearly independent.
- Given a set $B \in \mathcal{I}$ of linearly independent vectors, then any subset $A \subseteq B$ is also linearly independent. Hence, $\mathcal{I}$ is down-closed or "subclusive", under subsets. In other words,

$$A \subseteq B \text{ and } B \in \mathcal{I} \Rightarrow A \in \mathcal{I} \tag{5.96}$$

- maxInd: Inclusionwise maximal independent subsets (or bases) of any set $B \subseteq V$.

$$\text{maxInd}(B) \triangleq \{A \subseteq B : A \in \mathcal{I} \text{ and } \forall v \in B \setminus A, A \cup \{v\} \notin \mathcal{I}\} \tag{5.97}$$

- Given any set $B \subset V$ of vectors, all maximal (by set inclusion) subsets of linearly independent vectors are the same size. That is, for all $B \subseteq V$,

$$\forall A_1, A_2 \in \text{maxInd}(B), \quad |A_1| = |A_2| = \text{rank}(B) \tag{5.98}$$

From Matrix Rank $\rightarrow$ Matroid

- Let $\mathcal{I} = \{I_1, I_2, \dots\}$ be the set of sets as described above.
- Thus, for all $I \in \mathcal{I}$, the matrix rank function has the property

$$r(I) = |I| \tag{5.99}$$

and for any $B \notin \mathcal{I}$,

$$r(B) = \max\{|A| : A \subseteq B \text{ and } A \in \mathcal{I}\} < |B| \tag{5.100}$$

Since all maximally independent subsets of a set are the same size, the rank function is well defined.

Matroid

- Matroids abstract the notion of linear independence of a set of vectors to general algebraic properties.
- In a matroid, there is an underlying ground set, say E (or V), and a collection of subsets $\mathcal{I} = \{I_1, I_2, \dots\}$ of E that correspond to independent elements.
- There are many definitions of matroids that are mathematically equivalent, we'll see some of them here.

Independence System

Definition 5.6.1 (set system)

A (finite) ground set E and a set of subsets of E, $\emptyset \neq \mathcal{I} \subseteq 2^E$ is called a set system, notated $(E, \mathcal{I})$.

- Set systems can be arbitrarily complex since, as stated, there is no systematic method (besides exponential-cost exhaustive search) to determine if a given set $S \subseteq E$ has $S \in \mathcal{I}$.
- One useful property is "heredity." Namely, a set system is a hereditary set system if for any $A \subset B \in \mathcal{I}$, we have that $A \in \mathcal{I}$.

Independence System

Definition 5.6.2 (independence (or hereditary) system)

A set system $(V, \mathcal{I})$ is an independence system if

$$\emptyset \in \mathcal{I} \quad \text{(emptyset containing)} \tag{I1}$$

and

$$\forall I \in \mathcal{I}, J \subset I \Rightarrow J \in \mathcal{I} \quad \text{(subclusive)} \tag{I2}$$

- Property (I2) called "down monotone," "down closed," or "subclusive"
- Example: $E = \{1, 2, 3, 4\}$. With $\mathcal{I} = \{\emptyset, \{1\}, \{1, 2\}, \{1, 2, 4\}\}$.
- Then $(E, \mathcal{I})$ is a set system, but not an independence system since it is not down closed (i.e., we have $\{1, 2\} \in \mathcal{I}$ but not $\{2\} \in \mathcal{I}$).
- With $\mathcal{I} = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$, then $(E, \mathcal{I})$ is now an independence (hereditary) system.

Independence System

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{pmatrix} 0 & 0 & 1 & 1 & 2 & 1 & 3 & 1 \\ 0 & 1 & 1 & 0 & 2 & 0 & 2 & 4 \\ 1 & 1 & 1 & 0 & 0 & 3 & 1 & 5 \end{pmatrix} \end{array} = \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{pmatrix} | & | & | & | & | & | & | & | \\ x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 \\ | & | & | & | & | & | & | & | \end{pmatrix} \end{array} \quad (5.101)$$

- Given any set of linearly independent vectors A, any subset $B \subset A$ will also be linearly independent.
- Given any forest G_f that is an edge-induced sub-graph of a graph G, any sub-graph of G_f is also a forest.
- So these both constitute independence systems.

Matroid

Independent set definition of a matroid is perhaps most natural. Note, if $J \in \mathcal{I}$, then J is said to be an independent set.

Definition 5.6.3 (Matroid)

A set system $(E, \mathcal{I})$ is a Matroid if

(I1) $\emptyset \in \mathcal{I}$

(I2) $\forall I \in \mathcal{I}, J \subset I \Rightarrow J \in \mathcal{I}$ (down-closed or subclusive)

(I3) $\forall I, J \in \mathcal{I}$, with $|I| = |J| + 1$, then there exists $x \in I \setminus J$ such that $J \cup \{x\} \in \mathcal{I}$.

Why is (I1) is not redundant given (I2)? Because without (I1) could have a non-matroid where $\mathcal{I} = \{\}$.

On Matroids

- Abstract properties of linear dependence (Hassler Whitney, 1935), but already then found instances of objects with those properties not based on a matrix.
- Takeo Nakasawa, 1935, also early work.
- Forgotten for 20 years until mid 1950s.
- Matroids are powerful and flexible combinatorial objects.
- The rank function of a matroid is already a very powerful submodular function (perhaps all we need for many problems).
- Understanding matroids crucial for understanding submodularity.
- Matroid independent sets (i.e., A s.t. $r(A) = |A|$) are useful constraint set, and fast algorithms for submodular optimization subject to one (or more) matroid independence constraints exist.
- Crapo & Rota preferred the term "combinatorial geometry", or more specifically a "pregeometry" and said that pregeometries are "often described by the ineffably cacaphonic [sic] term 'matroid', which we prefer to avoid in favor of the term 'pregeometry'."

Matroid

Slight modification (non unit increment) that is equivalent.

Definition 5.6.4 (Matroid-II)

A set system $(E, \mathcal{I})$ is a Matroid if

(I1') $\emptyset \in \mathcal{I}$

(I2') $\forall I \in \mathcal{I}, J \subset I \Rightarrow J \in \mathcal{I}$ (down-closed or subclusive)

(I3') $\forall I, J \in \mathcal{I}$, with $|I| > |J|$, then there exists $x \in I \setminus J$ such that $J \cup \{x\} \in \mathcal{I}$

Note (I1)=(I1'), (I2)=(I2'), and we get (I3)≡(I3') using induction.

Matroids, independent sets, and bases

- Independent sets: Given a matroid $M = (E, \mathcal{I})$, a subset $A \subseteq E$ is called independent if $A \in \mathcal{I}$ and otherwise A is called dependent.
- A base of $U \subseteq E$: For $U \subseteq E$, a subset $B \subseteq U$ is called a base of U if B is inclusionwise maximally independent subset of U. That is, $B \in \mathcal{I}$ and there is no $Z \in \mathcal{I}$ with $B \subset Z \subseteq U$.
- A base of a matroid: If $U = E$, then a "base of E" is just called a base of the matroid M (this corresponds to a basis in a linear space, or a spanning forest in a graph, or a spanning tree in a connected graph).

Matroids - important property

Proposition 5.6.5

In a matroid $M = (E, \mathcal{I})$, for any $U \subseteq E(M)$, any two bases of U have the same size.

- In matrix terms, given a set of vectors U, all sets of independent vectors that span the space spanned by U have the same size.
- In fact, under (I1),(I2), this condition is equivalent to (I3). Exercise: show the following is equivalent to the above.

Definition 5.6.6 (Matroid)

A set system $(V, \mathcal{I})$ is a Matroid if

(I1') $\emptyset \in \mathcal{I}$ (emptyset containing)

(I2') $\forall I \in \mathcal{I}, J \subset I \Rightarrow J \in \mathcal{I}$ (down-closed or subclusive)

(I3') $\forall X \subseteq V$, and $I_1, I_2 \in \text{maxInd}(X)$, we have $|I_1| = |I_2|$ (all maximally independent subsets of X have the same size).

Matroids - rank

- Thus, in any matroid $M = (E, \mathcal{I})$, $\forall U \subseteq E(M)$, any two bases of U have the same size.
- The common size of all the bases of U is called the rank of U, denoted $r_M(U)$ or just $r(U)$ when the matroid in equation is unambiguous.
- $r(E) = r_{(E,\mathcal{I})}$ is the rank of the matroid, and is the common size of all the bases of the matroid.
- We can a bit more formally define the rank function this way.

Definition 5.6.7 (matroid rank function)

The rank function of a matroid is a function $r : 2^E \to \mathbb{Z}_+$ defined by

$$r(A) = \max\{|X| : X \subseteq A, X \in \mathcal{I}\} = \max_{X \in \mathcal{I}} |A \cap X| \tag{5.102}$$

- From the above, we immediately see that $r(A) \leq |A|$.
- Moreover, if $r(A) = |A|$, then $A \in \mathcal{I}$, meaning A is independent (in this case, A is a self base).

Matroids, other definitions using matroid rank $r : 2^V \to \mathbb{Z}_+$

Definition 5.6.8 (closed/flat/subspace)

A subset $A \subseteq E$ is closed (equivalently, a flat or a subspace) of matroid M if for all $x \in E \setminus A$, $r(A \cup \{x\}) = r(A) + 1$.

Definition: A hyperplane is a flat of rank $r(M) - 1$.

Definition 5.6.9 (closure)

Given $A \subseteq E$, the closure (or span) of A, is defined by $\text{span}(A) = \{b \in E : r(A \cup \{b\}) = r(A)\}$.

Therefore, a closed set A has $\text{span}(A) = A$.

Definition 5.6.10 (circuit)

A subset $A \subseteq E$ is circuit or a cycle if it is an inclusionwise-minimal dependent set (i.e., if $r(A) < |A|$ and for any $a \in A$, $r(A \setminus \{a\}) = |A| - 1$).

Matroids by bases

In general, besides independent sets and rank functions, there are other equivalent ways to characterize matroids.

Theorem 5.6.11 (Matroid (by bases))

Let E be a set and $\mathcal{B}$ be a nonempty collection of subsets of E. Then the following are equivalent.

1. *$\mathcal{B}$ is the collection of bases of a matroid;*
2. *if $B, B' \in \mathcal{B}$, and $x \in B' \setminus B$, then $B' - x + y \in \mathcal{B}$ for some $y \in B \setminus B'$.*
3. *If $B, B' \in \mathcal{B}$, and $x \in B' \setminus B$, then $B - y + x \in \mathcal{B}$ for some $y \in B \setminus B'$.*

Properties 2 and 3 are called "exchange properties."
Proof here is omitted but think about this for a moment in terms of linear spaces and matrices, and (alternatively) spanning trees.

Matroids by circuits

A set is independent if and only if it contains no circuit. Therefore, it is not surprising that circuits can also characterize a matroid.

Theorem 5.6.12 (Matroid by circuits)

Let E be a set and $\mathcal{C}$ be a collection of subsets of E that satisfy the following three properties:

1. *(C1): $\emptyset \notin \mathcal{C}$*
2. *(C2): if $C_1, C_2 \in \mathcal{C}$ and $C_1 \subseteq C_2$, then $C_1 = C_2$.*
3. *(C3): if $C_1, C_2 \in \mathcal{C}$ with $C_1 \neq C_2$, and $e \in C_1 \cap C_2$, then there exists a $C_3 \in \mathcal{C}$ such that $C_3 \subseteq (C_1 \cup C_2) \setminus \{e\}$.*

Matroids by circuits

Several circuit definitions for matroids.

Theorem 5.6.13 (Matroid by circuits)

Let E be a set and $\mathcal{C}$ be a collection of nonempty subsets of E, such that no two sets in $\mathcal{C}$ are contained in each other. Then the following are equivalent.

1. *$\mathcal{C}$ is the collection of circuits of a matroid;*
2. *if $C, C' \in \mathcal{C}$, and $x \in C \cap C'$, then $(C \cup C') \setminus \{x\}$ contains a set in $\mathcal{C}$;*
3. *if $C, C' \in \mathcal{C}$, and $x \in C \cap C'$, and $y \in C \setminus C'$, then $(C \cup C') \setminus \{x\}$ contains a set in $\mathcal{C}$ containing y;*

Again, think about this for a moment in terms of linear spaces and matrices, and spanning trees.