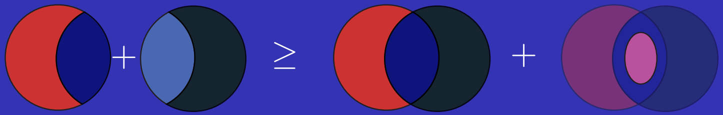


Deep Mathematical Properties of Submodularity with Applications to Machine Learning

Jeffrey A. Bilmes

Professor
Departments of Electrical Engineering
& Computer Science and Engineering
University of Washington, Seattle
http://melodi.ee.washington.edu/~bilmes

Thursday, December 5th, 2013

Outline

Outline

Where to get these slides

- Where to get these slides right now:

  `http://goo.gl/PSzuPv`

- QR Code:

A Dedication

This tutorial dedicated to Ben Taskar, and his family. RIP Ben.

Goals of the Tutorial

- Get an intuitive sense for submodular functions, should be able to apply them.
- Learn to recognize submodularity, or recognize when it might be useful.
- Learn to realize why submodularity can be useful in machine learning. Why is it worth your time to study it.
- Learn to realize when submodularity is inapplicable.

Submodularity

- Definition: given a finite ground set V, a function $f : 2^V \to \mathbb{R}$ is said to be submodular if

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B), \quad \forall A, B \subseteq V \tag{1}$$

Submodularity

- Definition: given a finite ground set V, a function $f : 2^V \to \mathbb{R}$ is said to be submodular if

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B), \quad \forall A, B \subseteq V \tag{1}$$

- The definition is the tip of the iceberg. This simple definition can lead to great mathematical and practical richness.

Submodularity

- Definition: given a finite ground set V, a function $f : 2^V \to \mathbb{R}$ is said to be submodular if

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B), \quad \forall A, B \subseteq V \tag{1}$$

- The definition is the tip of the iceberg. This simple definition can lead to great mathematical and practical richness.
- Goals of tutorial: will be very simple, an attempt to cover some important parts of the iceberg in 2 hours.

Submodularity

- Definition: given a finite ground set V, a function $f : 2^V \to \mathbb{R}$ is said to be submodular if

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B), \quad \forall A, B \subseteq V \tag{1}$$

- The definition is the tip of the iceberg. This simple definition can lead to great mathematical and practical richness.
- Goals of tutorial: will be very simple, an attempt to cover some important parts of the iceberg in 2 hours.
- The tutorial itself is the tip of the iceberg!

Submodularity

- Definition: given a finite ground set V, a function $f : 2^V \to \mathbb{R}$ is said to be submodular if

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B), \quad \forall A, B \subseteq V \tag{1}$$

- The definition is the tip of the iceberg. This simple definition can lead to great mathematical and practical richness.
- Goals of tutorial: will be very simple, an attempt to cover some important parts of the iceberg in 2 hours.
- The tutorial itself is the tip of the iceberg!
- One last goal: Let A be a set of tutorials on submodularity, and $f(A)$ the information provided by tutorials A. Our goal is to be a member of:

$$\operatorname*{argmax}_{v \in V \setminus B} f(B \cup \{v\}) \tag{2}$$

where B is the set of previous tutorials given on submodularity.

Outline

Sets and set functions

We are given a finite "ground" set of objects:

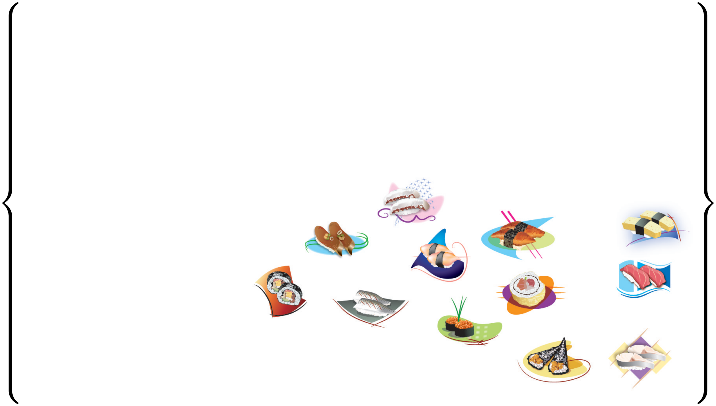
$V = \Bigg\{$  $\Bigg\}$

Also given a set function $f : 2^V \to \mathbb{R}$ that valuates subsets $A \subseteq V$.

Ex: $f(V) = 6$

Sets and set functions

Subset $A \subseteq V$ of objects:

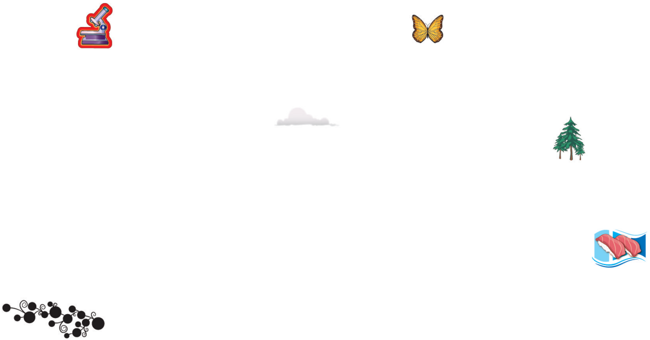
$A =$ 

Also given a set function $f : 2^V \to \mathbb{R}$ that valuates subsets $A \subseteq V$.

Ex: $f(A) = 1$

Sets and set functions

Subset $B \subseteq V$ of objects:

$B = \Big\{$  $\Big\}$

Also given a set function $f : 2^V \to \mathbb{R}$ that valuates subsets $A \subseteq V$.
Ex: $f(B) = 6$

Set functions are pseudo-Boolean functions

- Any set $A \subseteq V$ can be represented as a binary vector $x \in \{0,1\}^V$.

Set functions are pseudo-Boolean functions

- Any set $A \subseteq V$ can be represented as a binary vector $x \in \{0,1\}^V$.
- The characteristic vector of a set is given by $\mathbf{1}_A \in \{0,1\}^V$ where for all $v \in V$, we have:

$$\mathbf{1}_A(v) = \begin{cases} 1 & \text{if } v \in A \\ 0 & \textit{else} \end{cases} \tag{3}$$

Set functions are pseudo-Boolean functions

- Any set $A \subseteq V$ can be represented as a binary vector $x \in \{0,1\}^V$.
- The characteristic vector of a set is given by $\mathbf{1}_A \in \{0,1\}^V$ where for all $v \in V$, we have:

$$\mathbf{1}_A(v) = \begin{cases} 1 & \text{if } v \in A \\ 0 & \textit{else} \end{cases} \tag{3}$$

- It is sometimes useful to go back and forth between X and $x(X) \triangleq \mathbf{1}_X$.

Set functions are pseudo-Boolean functions

- Any set $A \subseteq V$ can be represented as a binary vector $x \in \{0,1\}^V$.
- The characteristic vector of a set is given by $\mathbf{1}_A \in \{0,1\}^V$ where for all $v \in V$, we have:

$$\mathbf{1}_A(v) = \begin{cases} 1 & \text{if } v \in A \\ 0 & \textit{else} \end{cases} \tag{3}$$

- It is sometimes useful to go back and forth between X and $x(X) \triangleq \mathbf{1}_X$.
- $f(x) : \{0,1\}^V \to \mathbb{R}$ is a pseudo-Boolean function, and submodular functions are a special case.

Two equivalent basic definitions

Definition (submodular)

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

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B) \tag{4}$$

Definition (submodular (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:

$$f(A \cup \{v\}) - f(A) \geq f(B \cup \{v\}) - f(B) \tag{5}$$

Outline

Example: Rank function of a matrix

- Given an $n \times m$ matrix, thought of as m column vectors:

$$\mathbf{X} = \begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & 4 & & m \end{array} \\ \begin{pmatrix} | & | & | & | & & | \\ x_1 & x_2 & x_3 & x_4 & \dots & x_m \\ | & | & | & | & & | \end{pmatrix} \end{array} \tag{6}$$

- Let set $V = \{1, 2, \dots, m\}$ be the set of column vector indices.
- For any subset of column vector indices $A \subseteq V$, let $r(A)$ be the rank of the column vectors indexed by A.
- Hence $r : 2^V \to \mathbb{Z}_+$ and $r(A)$ is the dimensionality of the vector space spanned by the set of vectors $\{x_a\}_{a \in A}$.
- Intuitively, $r(A)$ is the size of the largest set of independent vectors contained within the set of vectors indexed by A.

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

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

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

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

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

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

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

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

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

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

From Matrix Rank → Matroid

- So V is set of column vector indices of a matrix.

From Matrix Rank $\rightarrow$ Matroid

- So V is set of column vector indices of a matrix.
- Let $\mathcal{I}$ be a set of all subsets of V such that for any $I \in \mathcal{I}$, the vectors indexed by I are linearly independent.

From Matrix Rank $\rightarrow$ Matroid

- So V is set of column vector indices of a matrix.
- Let $\mathcal{I}$ 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.

From Matrix Rank $\rightarrow$ Matroid

- So V is set of column vector indices of a matrix.
- Let $\mathcal{I}$ 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.

From Matrix Rank $\rightarrow$ Matroid

- So V is set of column vector indices of a matrix.
- Let $\mathcal{I}$ 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,

$$A \subseteq B \text{ and } B \in \mathcal{I} \Rightarrow A \in \mathcal{I} \tag{7}$$

From Matrix Rank → Matroid

- So V is set of column vector indices of a matrix.
- Let $\mathcal{I}$ 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,

$$A \subseteq B \text{ and } B \in \mathcal{I} \Rightarrow A \in \mathcal{I} \tag{7}$$

- Inclusionwise maximal independent subsets (or bases) of B.

$$\text{maxInd}(B) \triangleq \{A \subseteq B : A \in \mathcal{I} \text{ and } \forall v \in B \setminus A, A \cup \{v\} \notin \mathcal{I}\} \tag{8}$$

From Matrix Rank → Matroid

- So V is set of column vector indices of a matrix.
- Let $\mathcal{I}$ 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,

$$A \subseteq B \text{ and } B \in \mathcal{I} \Rightarrow A \in \mathcal{I} \tag{7}$$

- Inclusionwise maximal independent subsets (or bases) of B.

$$\text{maxInd}(B) \triangleq \{A \subseteq B : A \in \mathcal{I} \text{ and } \forall v \in B \setminus A, A \cup \{v\} \notin \mathcal{I}\} \tag{8}$$

- 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 $B \subseteq V$,

$$\forall A_1, A_2 \in \text{maxInd}(B), \quad |A_1| = |A_2| \tag{9}$$

From Matrix Rank → Matroid

- Thus, for all $I \in \mathcal{I}$, the matrix rank function has the property

$$r(I) = |I| \tag{10}$$

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

$$r(B) = \max\{|A| : A \subseteq B \text{ and } A \in \mathcal{I}\} \leq |B| \tag{11}$$

Independence System

Definition (set system)

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

Definition (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

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

Independence System

Definition (set system)

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

Definition (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

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

- Ex: $V = \{1, 2, 3, 4\}$, $\mathcal{I} = \{\emptyset, \{1\}, \{1, 2\}, \{1, 2, 4\}\}$. $(V, \mathcal{I})$ is a set system, but not an independence system.

Independence System

Definition (set system)

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

Definition (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

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

- Ex: $V = \{1, 2, 3, 4\}$, $\mathcal{I} = \{\emptyset, \{1\}, \{1, 2\}, \{1, 2, 4\}\}$. $(V, \mathcal{I})$ is a set system, but not an independence system.
- If $\mathcal{I} = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$, then $(V, \mathcal{I})$ is independence system.

Matroids, many equivalent definitions

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

Definition (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).

Linear (or Matric) Matroid

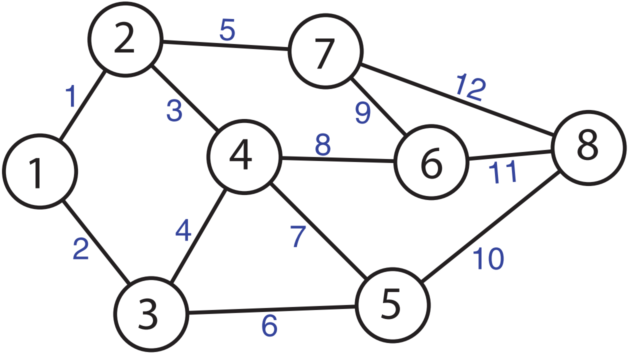
- Let $\mathbf{X}$ be an $n \times m$ matrix and $V = \{1, \ldots, m\}$
- Let $\mathcal{I}$ consists of subsets of V such that if $A \in \mathcal{I}$, and $A = \{a_1, a_2, \ldots, a_k\}$ then the vectors $x_{a_1}, x_{a_2}, \ldots, x_{a_k}$ are linearly independent.
- The rank function is just the rank of the space spanned by the corresponding set of vectors.
- A base of a matroid is maximally independent set. So a base of this matroid is a set of rank V independent vectors.

Cycle Matroid of a graph, or Graphic Matroids

- Let $G = (V, E)$ be a graph. Consider $(E, \mathcal{I})$ where the edges of the graph E are the ground set and $A \in \mathcal{I}$ if the edge-induced graph $G(V, A)$ by A does not contain any cycle.
- Then $M = (E, \mathcal{I})$ is a matroid.
- $\mathcal{I}$ contains all forests and trees.
- Bases are spanning forests (spanning trees if G is connected).
- Rank function $r(A)$ is the size of the largest spanning forest contained in $G(V, A)$.

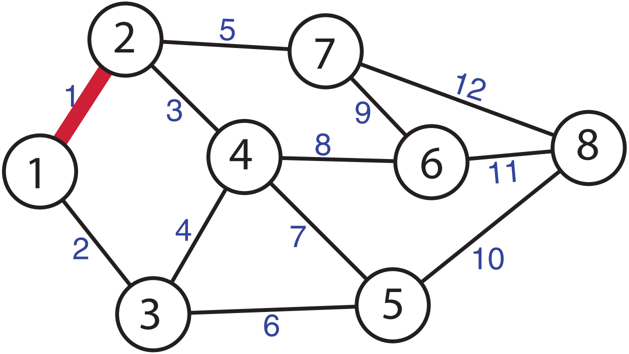
Example: graphic matroid

- A graph defines a matroid on edge sets, independent sets are those without a cycle.



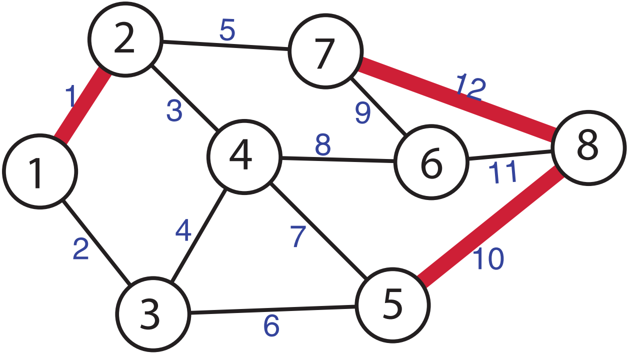
Example: graphic matroid

- A graph defines a matroid on edge sets, independent sets are those without a cycle.



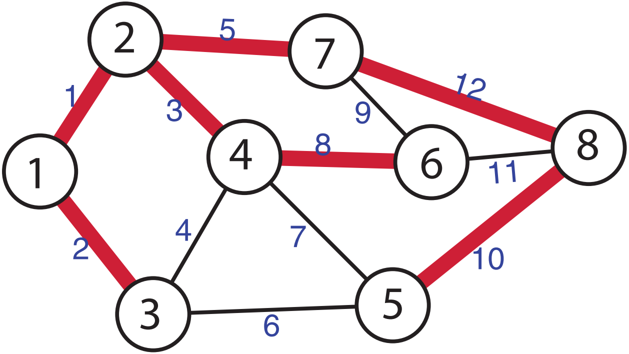
Example: graphic matroid

- A graph defines a matroid on edge sets, independent sets are those without a cycle.



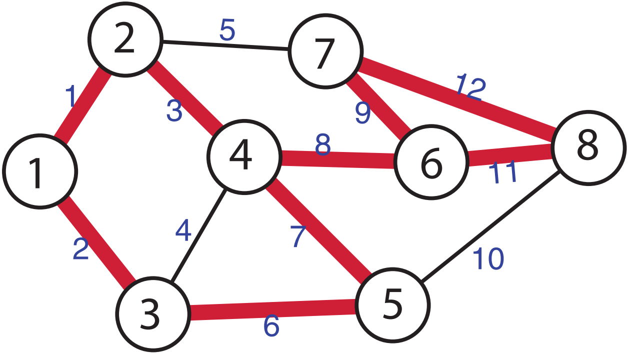
Example: graphic matroid

- A graph defines a matroid on edge sets, independent sets are those without a cycle.



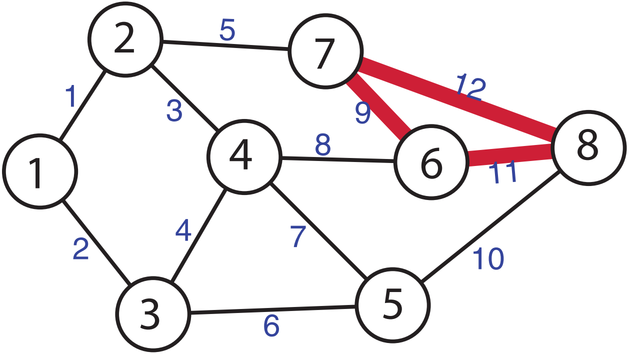
Example: graphic matroid

- A graph defines a matroid on edge sets, independent sets are those without a cycle.



Example: graphic matroid

- A graph defines a matroid on edge sets, independent sets are those without a cycle.



Partition Matroid

Ground set of objects, $V = \Big\{$

$\Big\}$

Partition Matroid

Partition of V into six blocks, $V_1, V_2, \ldots, V_6$

Partition Matroid

Limit associated with each block, $\{k_1, k_2, \ldots, k_6\}$

Partition Matroid

Independent subset but not maximally independent.

Partition Matroid

Maximally independent subset, what is called a base.

Partition Matroid

Not independent since over limit in set six.

Partition Matroid

- Let $V = V_1 \cup V_2 \cup \dots \cup V_\ell$ be a partition of V into disjoint sets (disjoint union). Define a set of subsets of V as

$$\mathcal{I} = \{X \subseteq V : |X \cap V_i| \leq k_i \text{ for all } i = 1, \dots, \ell\}. \tag{12}$$

  where $k_1, \dots, k_\ell$ are fixed parameters, $k_i \geq 0$. Then $M = (V, \mathcal{I})$ is a matroid.
- A partition matroids rank function is:

$$r(A) = \sum_{i=1}^{\ell} \min(|A \cap V_i|, k_i) \tag{13}$$

Matroids - rank and submodularity

We can a bit more formally define the rank function this way.

Definition

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

$$r(A) = \max \{|X| : X \subseteq A, X \in \mathcal{I}\} = \max_{X \in \mathcal{I}} |A \cap X| \tag{14}$$

Matroids - rank and submodularity

We can a bit more formally define the rank function this way.

Definition

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

$$r(A) = \max \{|X| : X \subseteq A, X \in \mathcal{I}\} = \max_{X \in \mathcal{I}} |A \cap X| \tag{14}$$

- From the above, we immediately see that $r(A) \leq |A|$.

Matroids - rank and submodularity

We can a bit more formally define the rank function this way.

Definition

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

$$r(A) = \max \{|X| : X \subseteq A, X \in \mathcal{I}\} = \max_{X \in \mathcal{I}} |A \cap X| \tag{14}$$

- 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

Matroids - rank and submodularity

We can a bit more formally define the rank function this way.

Definition

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

$$r(A) = \max\{|X| : X \subseteq A, X \in \mathcal{I}\} = \max_{X \in \mathcal{I}} |A \cap X| \tag{14}$$

- 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

Lemma

The rank function $r : 2^V \to \mathbb{Z}_+$ of a matroid is submodular, that is $r(A) + r(B) \geq r(A \cup B) + r(A \cap B)$

Matroids via rank

In fact, we can use the rank of a matroid for its definition.

Theorem (Matroid from rank)

Let V be a set and let $r : 2^V \to \mathbb{Z}_+$ be a function. Then $r(\cdot)$ defines a matroid with r being its rank function if and only if for all $A, B \subseteq V$:

(R1) $\forall A \subseteq V \;\; 0 \leq r(A) \leq |A|$ *(non-negative cardinality bounded)*

(R2) $r(A) \leq r(B)$ *whenever* $A \subseteq B \subseteq V$ *(monotone non-decreasing)*

(R3) $r(A \cup B) + r(A \cap B) \leq r(A) + r(B)$ *for all* $A, B \subseteq V$ *(submodular)*

- Unit increment (if $r(A) = k$, then either $r(A \cup \{v\}) = k$ or $r(A \cup \{v\}) = k + 1$) follows from the above.
- So submodularity and non-negative monotone non-decreasing, and unit increase are necessary and sufficient to define the matroid.

Modular functions, vectors

- Any set function $m : 2^V \to \mathbb{R}$ whose valuations, for all $A \subseteq V$, take the form

$$m(A) = \sum_{a \in A} m(a) \tag{15}$$

is called modular and normalized (meaning $m(\emptyset) = 0$).

Modular functions, vectors

- Any set function $m : 2^V \to \mathbb{R}$ whose valuations, for all $A \subseteq V$, take the form

$$m(A) = \sum_{a \in A} m(a) \tag{15}$$

  is called modular and normalized (meaning $m(\emptyset) = 0$).
- Any normalized modular function is identical to a vector $m \in \mathbb{R}^V$.

Modular functions, vectors

- Any set function $m : 2^V \to \mathbb{R}$ whose valuations, for all $A \subseteq V$, take the form

$$m(A) = \sum_{a \in A} m(a) \tag{15}$$

is called modular and normalized (meaning $m(\emptyset) = 0$).

- Any normalized modular function is identical to a vector $m \in \mathbb{R}^V$.
- Modular functions are submodular since $m(A) + m(B) \geq m(A \cup B) + m(A \cap B)$.

Modular functions, vectors

- Any set function $m : 2^V \to \mathbb{R}$ whose valuations, for all $A \subseteq V$, take the form

$$m(A) = \sum_{a \in A} m(a) \tag{15}$$

  is called modular and normalized (meaning $m(\emptyset) = 0$).
- Any normalized modular function is identical to a vector $m \in \mathbb{R}^V$.
- Modular functions are submodular since $m(A) + m(B) \geq m(A \cup B) + m(A \cap B)$.
- Modular functions are also supermodular since $m(A) + m(B) \leq m(A \cup B) + m(A \cap B)$.

Modular functions, vectors

- Any set function $m : 2^V \to \mathbb{R}$ whose valuations, for all $A \subseteq V$, take the form

$$m(A) = \sum_{a \in A} m(a) \tag{15}$$

  is called modular and normalized (meaning $m(\emptyset) = 0$).
- Any normalized modular function is identical to a vector $m \in \mathbb{R}^V$.
- Modular functions are submodular since $m(A) + m(B) \geq m(A \cup B) + m(A \cap B)$.
- Modular functions are also supermodular since $m(A) + m(B) \leq m(A \cup B) + m(A \cap B)$.
- Hence, the characteristic vector $\mathbf{1}_A$ of a set is modular.

Modular functions, vectors

- Any set function $m : 2^V \to \mathbb{R}$ whose valuations, for all $A \subseteq V$, take the form

$$m(A) = \sum_{a \in A} m(a) \tag{15}$$

  is called modular and normalized (meaning $m(\emptyset) = 0$).
- Any normalized modular function is identical to a vector $m \in \mathbb{R}^V$.
- Modular functions are submodular since $m(A) + m(B) \geq m(A \cup B) + m(A \cap B)$.
- Modular functions are also supermodular since $m(A) + m(B) \leq m(A \cup B) + m(A \cap B)$.
- Hence, the characteristic vector $\mathbf{1}_A$ of a set is modular.
- Modular functions are often called additive or linear.

Matroid and the greedy algorithm

- Let $(V, \mathcal{I})$ be an independence system, and we are given a non-negative modular weight function $w : V \to \mathbb{R}_+$.

Matroid and the greedy algorithm

- Let $(V, \mathcal{I})$ be an independence system, and we are given a non-negative modular weight function $w : V \to \mathbb{R}_+$.

Algorithm 1: The Matroid Greedy Algorithm

1 Set $X \leftarrow \emptyset$;
2 **while** $\exists v \in V \setminus X$ s.t. $X \cup \{v\} \in \mathcal{I}$ **do**
3 $\quad v \in \operatorname{argmax}\{w(v) : v \in V \setminus X,\ X \cup \{v\} \in \mathcal{I}\}$;
4 $\quad X \leftarrow X \cup \{v\}$;

Matroid and the greedy algorithm

- Let $(V, \mathcal{I})$ be an independence system, and we are given a non-negative modular weight function $w : V \to \mathbb{R}_+$.

Algorithm 1: The Matroid Greedy Algorithm

Set $X \leftarrow \emptyset$;
while $\underline{\exists v \in V \setminus X \text{ s.t. } X \cup \{v\} \in \mathcal{I}}$ **do**
  $v \in \operatorname{argmax} \{w(v) : v \in V \setminus X, \; X \cup \{v\} \in \mathcal{I}\}$;
  $X \leftarrow X \cup \{v\}$;

- Same as sorting items by decreasing weight w, and then choosing items in that order that retain independence.

Matroid and the greedy algorithm

- Let $(V, \mathcal{I})$ be an independence system, and we are given a non-negative modular weight function $w : V \to \mathbb{R}_+$.

Algorithm 1: The Matroid Greedy Algorithm

```
Set X ← ∅ ;
while ∃v ∈ V \ X s.t. X ∪ {v} ∈ I do
    v ∈ argmax {w(v) : v ∈ V \ X, X ∪ {v} ∈ I} ;
    X ← X ∪ {v} ;
```

- Same as sorting items by decreasing weight w, and then choosing items in that order that retain independence.

Theorem

Let $(V, \mathcal{I})$ be an independence system. Then the pair $(V, \mathcal{I})$ is a matroid if and only if for each weight function $w \in \mathcal{R}_+^V$, Algorithm 1 leads to a set $I \in \mathcal{I}$ of maximum weight $w(I)$.

Matroid Definitions Summary

Given an independence system, matroids are defined equivalently by any of the following:

- All maximally independent sets have the same size.
- A monotone non-decreasing submodular integral rank function with unit increments.
- The greedy algorithm achieves the maximum weight independent set for all weight functions.

Maximal points in a set

- Regarding sets, a subset X of S is a maximal subset of S possessing a given property P if X possesses property P and no set properly containing X (i.e., any $X' \supset X$ with $X' \setminus X \subseteq V \setminus X$) possesses P.

Maximal points in a set

- Regarding sets, a subset X of S is a maximal subset of S possessing a given property P if X possesses property P and no set properly containing X (i.e., any $X' \supset X$ with $X' \setminus X \subseteq V \setminus X$) possesses P.
- Given any set $P \subseteq \mathbb{R}^V$, we say that a vector x is maximal within P if it is the case that for any $\epsilon > 0$, and for all $v \in V$, we have that

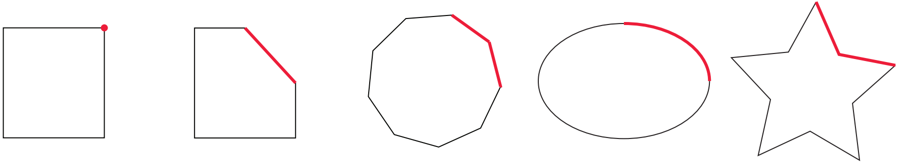
$$x + \epsilon \mathbf{1}_v \notin P \tag{16}$$

Maximal points in a set

- Regarding sets, a subset X of S is a maximal subset of S possessing a given property P if X possesses property P and no set properly containing X (i.e., any $X' \supset X$ with $X' \setminus X \subseteq V \setminus X$) possesses P.
- Given any set $P \subseteq \mathbb{R}^V$, we say that a vector x is maximal within P if it is the case that for any $\epsilon > 0$, and for all $v \in V$, we have that

$$x + \epsilon \mathbf{1}_v \notin P \tag{16}$$

- Examples of maximal regions (in red)

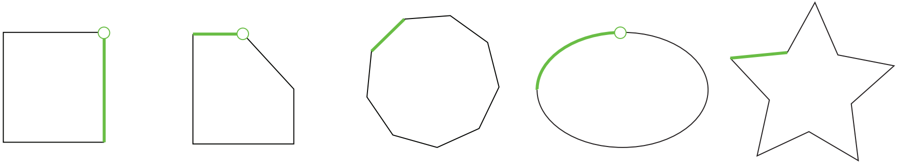


Maximal points in a set

- Regarding sets, a subset X of S is a maximal subset of S possessing a given property P if X possesses property P and no set properly containing X (i.e., any $X' \supset X$ with $X' \setminus X \subseteq V \setminus X$) possesses P.
- Given any set $P \subseteq \mathbb{R}^V$, we say that a vector x is maximal within P if it is the case that for any $\epsilon > 0$, and for all $v \in V$, we have that

$$x + \epsilon \mathbf{1}_v \notin P \tag{16}$$

- Examples of non-maximal regions (in green)



P-basis of x given compact set $P \subseteq \mathbb{R}_+^V$

Definition (subvector)

y is a subvector of x if $y \leq x$ (meaning $y(v) \leq x(v)$ for all $v \in V$).

P-basis of x given compact set $P \subseteq \mathbb{R}_+^V$

Definition (subvector)

y is a subvector of x if $y \leq x$ (meaning $y(v) \leq x(v)$ for all $v \in V$).

Definition (P-basis)

Given a compact set $P \subseteq \mathcal{R}_+^V$, for any $x \in \mathbb{R}_+^V$, a subvector y of x is called a P-basis of x if y maximal in P.
In other words, y is a P-basis of x if y is a maximal P-contained subvector of x.

Here, by y being "maximal", we mean that there exists no $z > y$ (more precisely, no $z \geq y + \epsilon \mathbf{1}_v$ for some $v \in V$ and $\epsilon > 0$) having the properties of y (the properties of y being: in P, and a subvector of x).

The “rank” of a vector

- Recall the definition of rank from a matroid $M = (V, \mathcal{I})$.

$$\text{rank}(A) = \max \{|I| : I \subseteq A, I \in \mathcal{I}\} \tag{17}$$

The "rank" of a vector

- Recall the definition of rank from a matroid $M = (V, \mathcal{I})$.

$$\text{rank}(A) = \max\{|I| : I \subseteq A, I \in \mathcal{I}\} \tag{17}$$

- vector rank: Given a compact set $P \subseteq \mathcal{R}_+^V$, we can define a form of "vector rank" relative to this P in the following way: Given an $x \in \mathbb{R}^V$, we define the vector rank, relative to P, as:

$$\text{rank}(x) = \max(y(V) : y \leq x, y \in P) \tag{18}$$

where $y \leq x$ is componentwise inequality ($y_i \leq x_i, \forall i$).

The "rank" of a vector

- Recall the definition of rank from a matroid $M = (V, \mathcal{I})$.

$$\text{rank}(A) = \max \{|I| : I \subseteq A, I \in \mathcal{I}\} \tag{17}$$

- vector rank: Given a compact set $P \subseteq \mathcal{R}_+^V$, we can define a form of "vector rank" relative to this P in the following way: Given an $x \in \mathbb{R}^V$, we define the vector rank, relative to P, as:

$$\text{rank}(x) = \max \left(y(V) : y \leq x, y \in P\right) \tag{18}$$

where $y \leq x$ is componentwise inequality ($y_i \leq x_i, \forall i$).

- If $\mathcal{B}_x$ is the set of P-bases of x, than $\text{rank}(x) = \max_{y \in \mathcal{B}_x} y(V)$.

The "rank" of a vector

- Recall the definition of rank from a matroid $M = (V, \mathcal{I})$.

$$\text{rank}(A) = \max\{|I| : I \subseteq A, I \in \mathcal{I}\} \tag{17}$$

- vector rank: Given a compact set $P \subseteq \mathcal{R}_+^V$, we can define a form of "vector rank" relative to this P in the following way: Given an $x \in \mathbb{R}^V$, we define the vector rank, relative to P, as:

$$\text{rank}(x) = \max\left(y(V) : y \leq x, y \in P\right) \tag{18}$$

  where $y \leq x$ is componentwise inequality $(y_i \leq x_i, \forall i)$.
- If $\mathcal{B}_x$ is the set of P-bases of x, than $\text{rank}(x) = \max_{y \in \mathcal{B}_x} y(V)$.
- If $x \in P$, then $\text{rank}(x) = x(V)$ (x is its own unique P-basis).

The "rank" of a vector

- Recall the definition of rank from a matroid $M = (V, \mathcal{I})$.

$$\text{rank}(A) = \max\{|I| : I \subseteq A, I \in \mathcal{I}\} \tag{17}$$

- vector rank: Given a compact set $P \subseteq \mathcal{R}_+^V$, we can define a form of "vector rank" relative to this P in the following way: Given an $x \in \mathbb{R}^V$, we define the vector rank, relative to P, as:

$$\text{rank}(x) = \max\left(y(V) : y \leq x, y \in P\right) \tag{18}$$

  where $y \leq x$ is componentwise inequality ($y_i \leq x_i, \forall i$).
- If $\mathcal{B}_x$ is the set of P-bases of x, than $\text{rank}(x) = \max_{y \in \mathcal{B}_x} y(V)$.
- If $x \in P$, then $\text{rank}(x) = x(V)$ (x is its own unique P-basis).
- In general, this might be hard to compute and/or have ill-defined properties. We next look at an object that restrains and cultivates this form of rank.

Polymatroidal polyhedron (or a "polymatroid")

Definition (polymatroid)

A polymatroid is a compact set $P \subseteq \mathbb{R}_+^V$ satisfying

1. $0 \in P$
2. If $y \leq x \in P$ then $y \in P$ (called down monotone).
3. For every $x \in \mathbb{R}_+^V$, any maximal vector $y \in P$ with $y \leq x$ (i.e., any P-basis of x), has the same component sum $y(V)$

Polymatroidal polyhedron (or a "polymatroid")

Definition (polymatroid)

A polymatroid is a compact set $P \subseteq \mathbb{R}_+^V$ satisfying

1. $0 \in P$
2. If $y \leq x \in P$ then $y \in P$ (called down monotone).
3. For every $x \in \mathbb{R}_+^V$, any maximal vector $y \in P$ with $y \leq x$ (i.e., any P-basis of x), has the same component sum $y(V)$

- Condition 3 restated: That is for any two distinct maximal vectors $y^1, y^2 \in P$, with $y^1 \leq x$ & $y^2 \leq x$, with $y^1 \neq y^2$, we must have $y^1(V) = y^2(V)$.

Polymatroidal polyhedron (or a "polymatroid")

Definition (polymatroid)

A polymatroid is a compact set $P \subseteq \mathbb{R}_+^V$ satisfying

1. $0 \in P$
2. If $y \leq x \in P$ then $y \in P$ (called down monotone).
3. For every $x \in \mathbb{R}_+^V$, any maximal vector $y \in P$ with $y \leq x$ (i.e., any P-basis of x), has the same component sum $y(V)$

- Condition 3 restated: That is for any two distinct maximal vectors $y^1, y^2 \in P$, with $y^1 \leq x$ & $y^2 \leq x$, with $y^1 \neq y^2$, we must have $y^1(V) = y^2(V)$.
- Condition 3 restated (again): For every vector $x \in \mathbb{R}_+^V$, every maximal independent subvector y of x has the same component sum $y(V) = \text{rank}(x)$.

Polymatroidal polyhedron (or a "polymatroid")

Definition (polymatroid)

A polymatroid is a compact set $P \subseteq \mathbb{R}_+^V$ satisfying

1. $0 \in P$
2. If $y \leq x \in P$ then $y \in P$ (called down monotone).
3. For every $x \in \mathbb{R}_+^V$, any maximal vector $y \in P$ with $y \leq x$ (i.e., any P-basis of x), has the same component sum $y(V)$

- Condition 3 restated: That is for any two distinct maximal vectors $y^1, y^2 \in P$, with $y^1 \leq x$ & $y^2 \leq x$, with $y^1 \neq y^2$, we must have $y^1(V) = y^2(V)$.
- Condition 3 restated (again): For every vector $x \in \mathbb{R}_+^V$, every maximal independent subvector y of x has the same component sum $y(V) = \text{rank}(x)$.
- Condition 3 restated (yet again): All P-bases of x have the same component sum.

Polymatroidal polyhedron (or a "polymatroid")

Definition (polymatroid)

A polymatroid is a compact set $P \subseteq \mathbb{R}_+^V$ satisfying

1. $0 \in P$
2. If $y \leq x \in P$ then $y \in P$ (called down monotone).
3. For every $x \in \mathbb{R}_+^V$, any maximal vector $y \in P$ with $y \leq x$ (i.e., any P-basis of x), has the same component sum $y(V)$

- Vectors within P (i.e., any $y \in P$) are called independent, and any vector outside of P is called dependent.

Polymatroidal polyhedron (or a "polymatroid")

Definition (polymatroid)

A polymatroid is a compact set $P \subseteq \mathbb{R}_+^V$ satisfying

1. $0 \in P$
2. If $y \leq x \in P$ then $y \in P$ (called down monotone).
3. For every $x \in \mathbb{R}_+^V$, any maximal vector $y \in P$ with $y \leq x$ (i.e., any P-basis of x), has the same component sum $y(V)$

- Vectors within P (i.e., any $y \in P$) are called independent, and any vector outside of P is called dependent.
- Since all P-bases of x have the same component sum, if $\mathcal{B}_x$ is the set of P-bases of x, than $\text{rank}(x) = y(V)$ for any $y \in \mathcal{B}_x$.

Matroid and Polymatroid: side-by-side

A Matroid is:

A Polymatroid is:

Matroid and Polymatroid: side-by-side

A Matroid is:

1. a set system $(V, \mathcal{I})$

A Polymatroid is:

1. a compact set $P \subseteq \mathbb{R}_+^V$

Matroid and Polymatroid: side-by-side

A Matroid is:

1. a set system $(V, \mathcal{I})$
2. empty-set containing $\emptyset \in \mathcal{I}$

A Polymatroid is:

1. a compact set $P \subseteq \mathbb{R}_+^V$
2. zero containing, $\mathbf{0} \in P$

Matroid and Polymatroid: side-by-side

A Matroid is:

1. a set system $(V, \mathcal{I})$
2. empty-set containing $\emptyset \in \mathcal{I}$
3. subclusive, down closed, $\emptyset \subseteq I' \subseteq I \in \mathcal{I} \Rightarrow I' \in \mathcal{I}$.

A Polymatroid is:

1. a compact set $P \subseteq \mathbb{R}_+^V$
2. zero containing, $\mathbf{0} \in P$
3. down monotone, $0 \leq y \leq x \in P \Rightarrow y \in P$

Matroid and Polymatroid: side-by-side

A Matroid is:

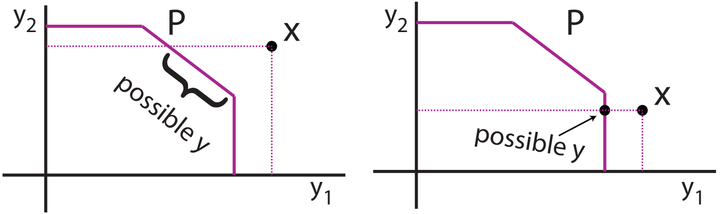
1. a set system $(V, \mathcal{I})$
2. empty-set containing $\emptyset \in \mathcal{I}$
3. subclusive, down closed, $\emptyset \subseteq I' \subseteq I \in \mathcal{I} \Rightarrow I' \in \mathcal{I}$.
4. any maximal set I in $\mathcal{I}$, bounded by another set A, has the same matroid rank (any maximal independent subset $I \subseteq A$ has same size $|I|$).

A Polymatroid is:

1. a compact set $P \subseteq \mathbb{R}_+^V$
2. zero containing, $\mathbf{0} \in P$
3. down monotone, $0 \leq y \leq x \in P \Rightarrow y \in P$
4. any maximal vector y in P, bounded by another vector x, has the same vector rank (any maximal independent subvector $y \leq x$ has same sum $y(V)$).

.

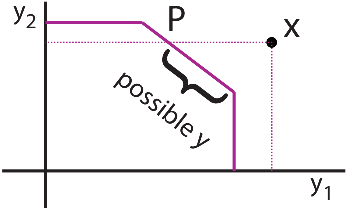
Polymatroidal polyhedron (or a "polymatroid")

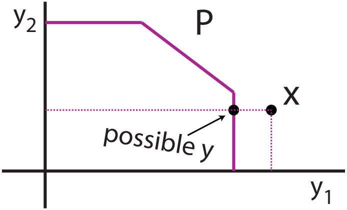


Left: $\exists$ multiple maximal $y \leq x$ Right: $\exists$ only one maximal $y \leq x$,

- Polymatroid condition here: $\forall$ maximal $y \in P$, with $y \leq x$ (which here means componentwise $y_1 \leq x_1$ and $y_2 \leq x_2$), we just have $y(V) = y_1 + y_2 =$ const.

Polymatroidal polyhedron (or a "polymatroid")

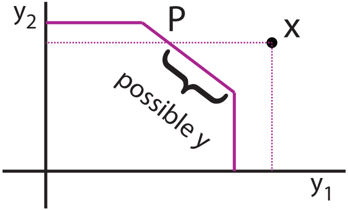


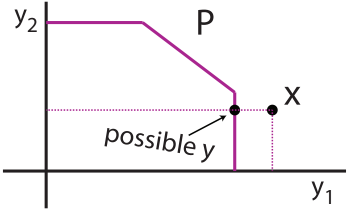


Left: $\exists$ multiple maximal $y \leq x$ Right: $\exists$ only one maximal $y \leq x$,

- Polymatroid condition here: $\forall$ maximal $y \in P$, with $y \leq x$ (which here means componentwise $y_1 \leq x_1$ and $y_2 \leq x_2$), we just have $y(V) = y_1 + y_2 = \text{const}$.
- On the left, we see there are multiple possible maximal $y \in P$ such that $y \leq x$. Each such y must have the same value $y(V)$.

Polymatroidal polyhedron (or a "polymatroid")

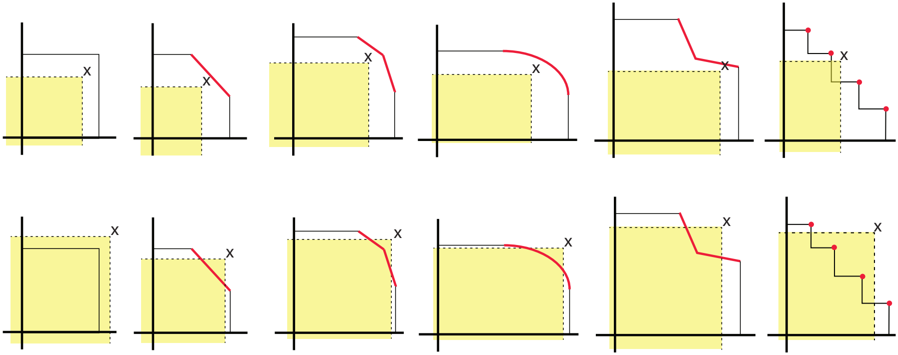




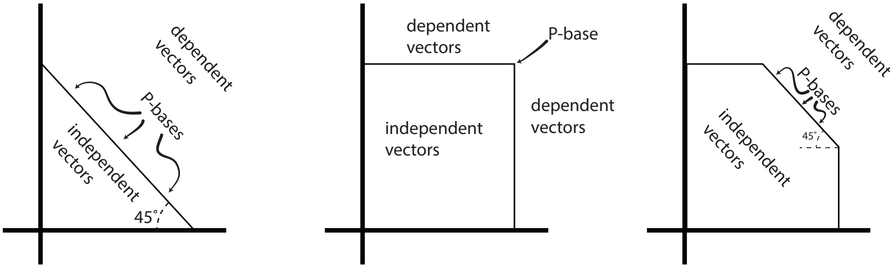
Left: $\exists$ multiple maximal $y \leq x$ Right: $\exists$ only one maximal $y \leq x$,

- Polymatroid condition here: $\forall$ maximal $y \in P$, with $y \leq x$ (which here means componentwise $y_1 \leq x_1$ and $y_2 \leq x_2$), we just have $y(V) = y_1 + y_2 = \text{const}$.
- On the left, we see there are multiple possible maximal $y \in P$ such that $y \leq x$. Each such y must have the same value $y(V)$.
- On the right, there is only one maximal $y \in P$. Since there is only one, the condition on the same value of $y(V), \forall y$ is vacuous.

Other examples: Polymatroid or not?



Some possible polymatroid forms in 2D



It appears that we have three possible forms of polymatroid in 2D, when neither of the elements $\{v_1, v_2\}$ are self-dependent.

1. On the left: full dependence between v_1 and v_2
2. In the middle: full independence between v_1 and v_2
3. On the right: partial independence between v_1 and v_2

- The P-bases (or single P-base in the middle case) are as indicated.
- Independent vectors are those within or on the boundary of the polytope. Dependent vectors are exterior to the polytope.
- The set of P-bases for a polytope is called the base polytope.

Polymatroid function and its polyhedron.

Definition

A polymatroid function is a real-valued function f defined on subsets of V which is normalized, non-decreasing, and submodular. That is:

1. $f(\emptyset) = 0$ (normalized)
2. $f(A) \leq f(B)$ for any $A \subseteq B \subseteq V$ (monotone non-decreasing)
3. $f(A \cup B) + f(A \cap B) \leq f(A) + f(B)$ for any $A, B \subseteq V$ (submodular)

We can define the polyhedron P_f^+ associated with a polymatroid function as follows

$$P_f^+ = \left\{ y \in \mathbb{R}_+^V : y(A) \leq f(A) \text{ for all } A \subseteq V \right\} \quad (19)$$

$$= \left\{ y \in \mathbb{R}^V : y \geq 0, y(A) \leq f(A) \text{ for all } A \subseteq V \right\} \quad (20)$$

A polymatroid function's polyhedron is a polymatroid.

Theorem (Edmonds, 1970)

Let f be a polymatroid function defined on subsets of V. For any $x \in \mathbb{R}_+^V$, and any P_f^+-basis $y^x \in \mathbb{R}_+^V$ of x, the component sum of y^x is

$$y^x(V) = rank(x) = \max\left(y(V) : y \leq x, y \in P_f^+\right) \\ = \min\left(x(A) + f(V \setminus A) : A \subseteq V\right) \quad (21)$$

As a consequence, P_f^+ is a polymatroid, since r.h.s. is constant w.r.t. y^x.

A polymatroid function's polyhedron is a polymatroid.

Theorem (Edmonds, 1970)

Let f be a polymatroid function defined on subsets of V. For any $x \in \mathbb{R}_+^V$, and any P_f^+-basis $y^x \in \mathbb{R}_+^V$ of x, the component sum of y^x is

$$y^x(V) = rank(x) = \max\left(y(V) : y \leq x, y \in P_f^+\right)$$
$$= \min\left(x(A) + f(V \setminus A) : A \subseteq V\right) \quad (21)$$

As a consequence, P_f^+ is a polymatroid, since r.h.s. is constant w.r.t. y^x.

With an appropriate choice of x, we can define/recover the submodular function from the polymatroid polyhedron via the following:

$$f(A) = \max\left\{y(A) : y \in P_f^+\right\} \quad (22)$$

There are many important consequences of this theorem (other than just P_f^+ is a polymatroid), regarding submodular function minimization.

A polymatroid is a polymatroid function's polytope

Thus, when f is a polymatroid function, P_f^+ is a polymatroid.

- Is it the case that, conversely, for any polymatroid P, there is an associated polymatroidal function f such that $P = P_f^+$?

A polymatroid is a polymatroid function's polytope

Thus, when f is a polymatroid function, P_f^+ is a polymatroid.

- Is it the case that, conversely, for any polymatroid P, there is an associated polymatroidal function f such that $P = P_f^+$?

Theorem

For any polymatroid P (compact subset of $\mathbb{R}_+^V$, zero containing, down-monotone, and $\forall x \in \mathbb{R}_+^V$ any maximal independent subvector $y \leq x$ has same component sum $y(V) = rank(x)$), there is a polymatroid function $f : 2^V \to \mathbb{R}$ (normalized, monotone non-decreasing, submodular) such that $P = P_f^+$ where $P_f^+ = \{x \in \mathbb{R}^V : x \geq 0, x(A) \leq f(A), \forall A \subseteq V\}$.

Next, a bit on rank(x)

- Recall that the matroid rank function is submodular.

Next, a bit on rank(x)

- Recall that the matroid rank function is submodular.
- For vectors $u, v \in \mathbb{R}^V$, let $u \vee v$ be the element-wise max (i.e., $(u \vee v)(i) = max(u(i), v(i))$), and $u \wedge v$ be elementwise min.

Next, a bit on rank(x)

- Recall that the matroid rank function is submodular.
- For vectors $u, v \in \mathbb{R}^V$, let $u \vee v$ be the element-wise max (i.e., $(u \vee v)(i) = max(u(i), v(i))$), and $u \wedge v$ be elementwise min.
- The polymatroid vector rank function rank(x) also satisfies a form of submodularity.

Next, a bit on rank(x)

- Recall that the matroid rank function is submodular.
- For vectors $u, v \in \mathbb{R}^V$, let $u \vee v$ be the element-wise max (i.e., $(u \vee v)(i) = max(u(i), v(i))$), and $u \wedge v$ be elementwise min.
- The polymatroid vector rank function rank(x) also satisfies a form of submodularity.

Theorem (vector rank and submodularity)

Let P be a polymatroid polytope. The vector rank function $rank : \mathbb{R}_+^V \to \mathbb{R}$ with $rank(x) = \max\left(y(V) : y \leq x, y \in P\right)$ satisfies, for all $u, v \in \mathbb{R}_+^V$

$$rank(u) + rank(v) \geq rank(u \vee v) + rank(u \wedge v) \tag{23}$$

Polymatroid from polymatroid function

- Recall, a matroid may be given as (V, r) where r is the rank function.
- We mention also that the term "polymatroid" is sometimes not used for the polytope itself, but instead but for the pair (V, f),
- Since (V, f) is equivalent to a polymatroid polytope, this is sensible.

Polymatroidal polyhedron and greedy

- Let $(V, \mathcal{I})$ be a set system and $w \in \mathbb{R}_+^V$ be a weight vector.

Polymatroidal polyhedron and greedy

- Let $(V, \mathcal{I})$ be a set system and $w \in \mathbb{R}_+^V$ be a weight vector.
- Recall greedy algorithm: Set $A = \emptyset$, and repeatedly choose $v \in V \setminus A$ such that $A \cup \{v\} \in \mathcal{I}$ with $w(v)$ as large as possible, stopping when no such v exists.

Polymatroidal polyhedron and greedy

- Let $(V, \mathcal{I})$ be a set system and $w \in \mathbb{R}_+^V$ be a weight vector.
- Recall greedy algorithm: Set $A = \emptyset$, and repeatedly choose $v \in V \setminus A$ such that $A \cup \{v\} \in \mathcal{I}$ with $w(v)$ as large as possible, stopping when no such v exists.
- For a matroid, we saw that set system $(V, \mathcal{I})$ is a matroid iff for each weight function $w \in \mathbb{R}_+^V$, the greedy algorithm leads to a set $I \in \mathcal{I}$ of maximum weight $w(I)$.

Polymatroidal polyhedron and greedy

- Let $(V, \mathcal{I})$ be a set system and $w \in \mathbb{R}_+^V$ be a weight vector.
- Recall greedy algorithm: Set $A = \emptyset$, and repeatedly choose $v \in V \setminus A$ such that $A \cup \{v\} \in \mathcal{I}$ with $w(v)$ as large as possible, stopping when no such v exists.
- For a matroid, we saw that set system $(V, \mathcal{I})$ is a matroid iff for each weight function $w \in \mathbb{R}_+^V$, the greedy algorithm leads to a set $I \in \mathcal{I}$ of maximum weight $w(I)$.
- Can we also characterize a polymatroid in this way?

Polymatroidal polyhedron and greedy

- Let $(V, \mathcal{I})$ be a set system and $w \in \mathbb{R}_+^V$ be a weight vector.
- Recall greedy algorithm: Set $A = \emptyset$, and repeatedly choose $v \in V \setminus A$ such that $A \cup \{v\} \in \mathcal{I}$ with $w(v)$ as large as possible, stopping when no such v exists.
- For a matroid, we saw that set system $(V, \mathcal{I})$ is a matroid iff for each weight function $w \in \mathbb{R}_+^V$, the greedy algorithm leads to a set $I \in \mathcal{I}$ of maximum weight $w(I)$.
- Can we also characterize a polymatroid in this way?
- That is, if we consider $\max \{wx : x \in P_f^+\}$, where P_f^+ represents the "independent vectors", is it the case that P_f^+ is a polymatroid iff greedy works for this maximization?

Chains of sets

- Ground element $V = \{1, 2, \ldots, n\}$ set of integers w.l.o.g.

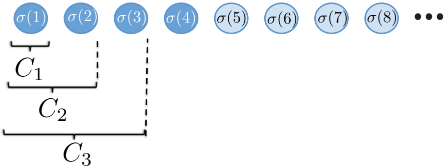
Chains of sets

- Ground element $V = \{1, 2, \ldots, n\}$ set of integers w.l.o.g.
- Given a permutation $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$ of the integers.

Chains of sets

- Ground element $V = \{1, 2, \ldots, n\}$ set of integers w.l.o.g.
- Given a permutation $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$ of the integers.
- From this we can form a chain of sets $\{C_i\}_i$ with $\emptyset = C_0 \subseteq C_1 \subseteq \cdots \subseteq C_n = V$ formed as:

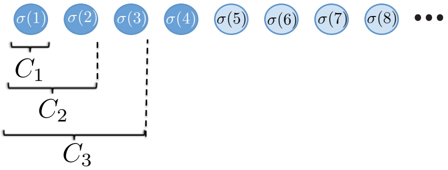
$$C_i = \{\sigma_1, \sigma_2, \ldots, \sigma_i\}, \quad \text{for } i = 1 \ldots n \tag{24}$$



Chains of sets

- Ground element $V = \{1, 2, \ldots, n\}$ set of integers w.l.o.g.
- Given a permutation $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$ of the integers.
- From this we can form a chain of sets $\{C_i\}_i$ with $\emptyset = C_0 \subseteq C_1 \subseteq \cdots \subseteq C_n = V$ formed as:

$$C_i = \{\sigma_1, \sigma_2, \ldots, \sigma_i\}, \quad \text{for } i = 1 \ldots n \tag{24}$$



- Can also form a chain from a vector $w \in \mathbb{R}^V$ sorted in descending order. Choose σ so that $w(\sigma_1) \geq w(\sigma_2) \geq \cdots \geq w(\sigma_n)$.

Gain

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

Gain

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

$$f(A \cup \{j\}) - f(A) \triangleq \rho_j(A) \tag{25}$$

$$\triangleq \rho_A(j) \tag{26}$$

$$\triangleq \nabla_j f(A) \tag{27}$$

$$\triangleq f(\{j\}|A) \tag{28}$$

$$\triangleq f(j|A) \tag{29}$$

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:

$$
\begin{aligned}
f(A \cup \{j\}) - f(A) &\triangleq \rho_j(A) && (25)\\
&\triangleq \rho_A(j) && (26)\\
&\triangleq \nabla_j f(A) && (27)\\
&\triangleq f(\{j\}|A) && (28)\\
&\triangleq f(j|A) && (29)
\end{aligned}
$$

- We'll use $f(j|A)$. Also, $f(A|B) = f(A \cup B) - f(B)$.

Gain

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

$$f(A \cup \{j\}) - f(A) \triangleq \rho_j(A) \tag{25}$$

$$\triangleq \rho_A(j) \tag{26}$$

$$\triangleq \nabla_j f(A) \tag{27}$$

$$\triangleq f(\{j\}|A) \tag{28}$$

$$\triangleq f(j|A) \tag{29}$$

- We'll use $f(j|A)$. Also, $f(A|B) = f(A \cup B) - f(B)$.
- 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).

Polymatroidal polyhedron and greedy

- Greedy solution to $\max(wx : x \in P_f)$

Polymatroidal polyhedron and greedy

- Greedy solution to $\max(wx : x \in P_f)$
- Sort elements of V w.r.t. w so that, w.l.o.g. $V = (v_1, v_2, \dots, v_m)$ with $w(v_1) \geq w(v_2) \geq \dots \geq w(v_m)$.

Polymatroidal polyhedron and greedy

- Greedy solution to $\max(wx : x \in P_f)$
- Sort elements of V w.r.t. w so that, w.l.o.g. $V = (v_1, v_2, \dots, v_m)$ with $w(v_1) \geq w(v_2) \geq \dots \geq w(v_m)$.
- Next, form chain of sets based on w sorted descended, giving:

$$V_i \stackrel{\text{def}}{=} \{v_1, v_2, \dots v_i\} \tag{30}$$

for $i = 0 \dots m$. Note $V_0 = \emptyset$, and $f(V_0) = 0$.

Polymatroidal polyhedron and greedy

- Greedy solution to $\max(wx : x \in P_f)$
- Sort elements of V w.r.t. w so that, w.l.o.g. $V = (v_1, v_2, \ldots, v_m)$ with $w(v_1) \geq w(v_2) \geq \cdots \geq w(v_m)$.
- Next, form chain of sets based on w sorted descended, giving:

$$V_i \stackrel{\text{def}}{=} \{v_1, v_2, \ldots v_i\} \tag{30}$$

for $i = 0 \ldots m$. Note $V_0 = \emptyset$, and $f(V_0) = 0$.

- The greedy solution is the vector $x \in \mathbb{R}_+^V$ with element $x(v_i)$ for $i = 1, \ldots, n$ defined as:

$$x(v_i) = f(V_i) - f(V_{i-1}) = f(v_i | V_{i-1}) \tag{31}$$

Polymatroidal polyhedron and greedy

- We have a result very similar to what we saw for matroids.

Theorem

Let $f : 2^V \to \mathbb{R}_+$ be a given set function, and P is a polytope in $\mathbb{R}^V_+$ of the form $P = \left\{x \in \mathbb{R}^V_+ : x(A) \leq f(A), \forall A \subseteq V\right\}$.
Then the greedy solution to the problem $\max(wx : x \in P)$ is optimal $\forall w$ iff f is monotone non-decreasing submodular (i.e., iff P is a polymatroid).

Polymatroid extreme points

Greedy does more than this. In fact, we have:

Theorem

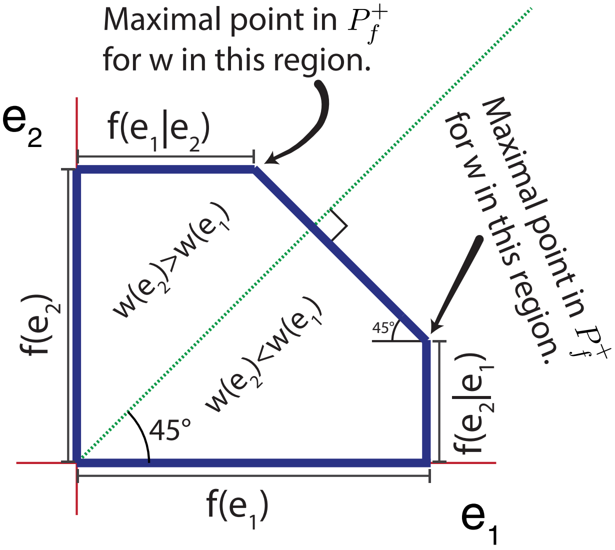
For a given ordering $V = (v_1, \dots, v_m)$ of V and a given V_i and x generated by V_i using the greedy procedure, then x is an extreme point of P_f

Corollary

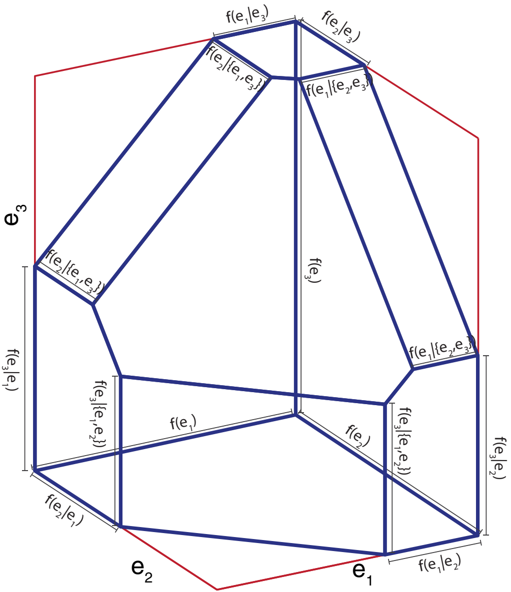
If x is an extreme point of P_f and $B \subseteq V$ is given such that $\{v \in V : x(v) \neq 0\} \subseteq B \subseteq \cup(A : x(A) = f(A))$, then x is generated using greedy by some ordering of B.

Intuition: why greedy works with polymatroids

- Given w, the goal is to find $x = (x(e_1), x(e_2))$ that maximizes $x^\intercal w = x(e_1)w(e_1) + x(e_2)w(e_2)$.
- If $w(e_2) > w(e_1)$ the upper extreme point indicated maximizes $x^\intercal w$ over $x \in P_f^+$.
- If $w(e_2) < w(e_1)$ the lower extreme point indicated maximizes $x^\intercal w$ over $x \in P_f^+$.



Polymatroid with labeled edge lengths



A polymatroid function's polyhedron vs. a polymatroid.

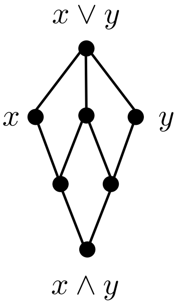
- Given these results, we can conclude that a polymatroid is really an extremely natural polyhedral generalization of a matroid. This was all realized by Jack Edmonds in the mid 1960s (and published in 1969 in his landmark paper "Submodular Functions, Matroids, and Certain Polyhedra").

- Jack Edmonds NIPS talk, 2011 `http://videolectures.net/nipsworkshops2011_edmonds_polymatroids/`

Outline

Submodular (or Upper-SemiModular) Lattices

The name "Submodular" comes from lattice theory, and refers to a property of the "height" function of an upper-semimodular lattice. Ex: consider the following lattice over 7 elements.



height

3

2

1

0

submodularity

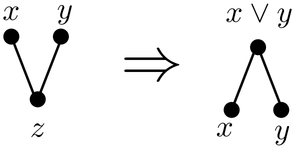
h(x)+h(y)

$>$ h($x \vee y$)

$+$ h($x \wedge y$)

$2 + 2 > 3 + 0$

- Such lattices require that for all x, y, z,



- The lattice is upper-semimodular (submodular), height function is submodular on the lattice.

Submodular Definitions

Definition (submodular)

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

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B) \tag{32}$$

- General submodular function, f need not be monotone, non-negative, nor normalized (i.e., $f(\emptyset)$ need not be $= 0$).

Normalized Submodular Function

- Given any submodular function $f : 2^V \to \mathbb{R}$, form a normalized variant $f' : 2^V \to \mathbb{R}$, with

$$f'(A) = f(A) - f(\emptyset) \tag{33}$$

- Then $f'(\emptyset) = 0$.
- This operation does not affect submodularity, or any minima or maxima
- We will assume that all functions in this tutorial are so normalized.

Submodular Polymatroidal Decomposition

- Given any arbitrary submodular function $f : 2^V \to \mathbb{R}$, consider the identity

$$f(A) = \underbrace{f(A) - m(A)}_{\bar{f}(A)} + m(A) = \bar{f}(A) + m(A) \tag{34}$$

for a modular function $m : 2^V \to \mathbb{R}$, where

$$m(a) = f(a|V \setminus \{a\}) \tag{35}$$

Submodular Polymatroidal Decomposition

- Given any arbitrary submodular function $f : 2^V \to \mathbb{R}$, consider the identity

$$f(A) = \underbrace{f(A) - m(A)}_{\bar{f}(A)} + m(A) = \bar{f}(A) + m(A) \tag{34}$$

for a modular function $m : 2^V \to \mathbb{R}$, where

$$m(a) = f(a|V \setminus \{a\}) \tag{35}$$

- Then $\bar{f}(A)$ is polymatroidal since $\bar{f}(\emptyset) = 0$ and for any a and A

$$\bar{f}(a|A) = f(a|A) - f(a|V \setminus \{a\}) \geq 0 \tag{36}$$

Totally Normalized

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

Totally Normalized

- $\bar{f}$ is called the totally normalized version of f
- polytope of $\bar{f}$ and f is the same shape, just shifted.

$$P_f = \left\{ x \in \mathbb{R}^V : x(A) \leq f(A), \forall A \subseteq V \right\} \tag{37}$$

$$= \left\{ x \in \mathbb{R}^V : x(A) \leq \bar{f}(A) + m(A), \forall A \subseteq V \right\} \tag{38}$$

Totally Normalized

- $\bar{f}$ is called the totally normalized version of f
- polytope of $\bar{f}$ and f is the same shape, just shifted.

$$P_f = \left\{ x \in \mathbb{R}^V : x(A) \le f(A), \forall A \subseteq V \right\} \tag{37}$$

$$= \left\{ x \in \mathbb{R}^V : x(A) \le \bar{f}(A) + m(A), \forall A \subseteq V \right\} \tag{38}$$

- m is like a unary score, $\bar{f}$ is where things interact . All of the real structure is in $\bar{f}$

Totally Normalized

- $\bar{f}$ is called the totally normalized version of f
- polytope of $\bar{f}$ and f is the same shape, just shifted.

$$P_f = \left\{ x \in \mathbb{R}^V : x(A) \leq f(A), \forall A \subseteq V \right\} \tag{37}$$
$$= \left\{ x \in \mathbb{R}^V : x(A) \leq \bar{f}(A) + m(A), \forall A \subseteq V \right\} \tag{38}$$

- m is like a unary score, $\bar{f}$ is where things interact . All of the real structure is in $\bar{f}$
- Hence, any submodular function is a sum of polymatroid and modular.

Telescoping Summation

- Given a chain set of sets $A_1 \subseteq A_2 \subseteq \cdots \subseteq A_r$

Telescoping Summation

- Given a chain set of sets $A_1 \subseteq A_2 \subseteq \cdots \subseteq A_r$
- Then the telescoping summation property of the gains is as follows:

$$\sum_{i=1}^{r-1} f(A_{i+1}|A_i) = \sum_{i=2}^{r} f(A_i) - \sum_{i=1}^{r-1} f(A_i) = f(A_r) - f(A_1) \tag{39}$$

Submodular Definitions

Theorem

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

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

if and only if

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

Submodular Definitions

Theorem

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

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

if and only if

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

Proof.

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

(DR)⇒(SC): Order $A \setminus B = \{v_1, v_2, \ldots, v_r\}$ arbitrarily. Then $f(v_i|A \cap B \cup \{v_1, v_2, \ldots, v_{i-1}\}) \geq f(v_1|B \cup \{v_1, v_2, \ldots, v_{i-1}\}),\ i \in [r-1]$
Applying telescoping summation to both sides, we get:

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

□

Many (Equivalent) Definitions of Submodularity

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

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

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

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

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

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

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

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

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

Basic ops: Sums, Restrictions, Conditioning

- Given submodular $f_1, f_2, \ldots, f_k$ each $\in 2^V \to \mathbb{R}$, then conic combinations are submodular. I.e.,

$$f(A) = \sum_{i=1}^{k} \alpha_i f_i(A) \tag{40}$$

where $\alpha_i \geq 0$.

Basic ops: Sums, Restrictions, Conditioning

- Given submodular $f_1, f_2, \ldots, f_k$ each $\in 2^V \to \mathbb{R}$, then conic combinations are submodular. I.e.,

$$f(A) = \sum_{i=1}^{k} \alpha_i f_i(A) \tag{40}$$

  where $\alpha_i \geq 0$.
- Restrictions: $f(A) = g(A \cap C)$ is submodular whenever g is, for all C.

Basic ops: Sums, Restrictions, Conditioning

- Given submodular $f_1, f_2, \ldots, f_k$ each $\in 2^V \to \mathbb{R}$, then conic combinations are submodular. I.e.,

$$f(A) = \sum_{i=1}^{k} \alpha_i f_i(A) \tag{40}$$

  where $\alpha_i \geq 0$.
- Restrictions: $f(A) = g(A \cap C)$ is submodular whenever g is, for all C.
- Conditioning: $f(A) = g(A \cup C) - f(C) = f(A|C)$ is submodular whenever g is for all C.

The "or" of two polymatroid functions

- Given two polymatroid functions f and g, suppose feasible A are defined as $\{A : f(A) \geq \alpha_f \text{ or } g(A) \geq \alpha_g\}$ for real α_f, α_g.

The "or" of two polymatroid functions

- Given two polymatroid functions f and g, suppose feasible A are defined as $\{A : f(A) \geq \alpha_f \text{ or } g(A) \geq \alpha_g\}$ for real α_f, α_g.
- This is identical to: $\{A : \bar{f}(A) = \alpha_f \text{ or } \bar{g}(A) = \alpha_g\}$ where $\bar{f}(A) = \min(f(A), \alpha_f)$ and $\bar{g}(A) = \min(f(A), \alpha_g)$

The "or" of two polymatroid functions

- Given two polymatroid functions f and g, suppose feasible A are defined as $\{A : f(A) \geq \alpha_f \text{ or } g(A) \geq \alpha_g\}$ for real α_f, α_g.
- This is identical to: $\{A : \bar{f}(A) = \alpha_f \text{ or } \bar{g}(A) = \alpha_g\}$ where $\bar{f}(A) = \min(f(A), \alpha_f)$ and $\bar{g}(A) = \min(f(A), \alpha_g)$
- Define: $h(A) = \bar{f}(A)\bar{g}(V) + \bar{f}(V)\bar{g}(A) - \bar{f}(A)\bar{g}(A)$.

The "or" of two polymatroid functions

- Given two polymatroid functions f and g, suppose feasible A are defined as $\{A : f(A) \geq \alpha_f \text{ or } g(A) \geq \alpha_g\}$ for real α_f, α_g.
- This is identical to: $\{A : \bar{f}(A) = \alpha_f \text{ or } \bar{g}(A) = \alpha_g\}$ where $\bar{f}(A) = \min(f(A), \alpha_f)$ and $\bar{g}(A) = \min(f(A), \alpha_g)$
- Define: $h(A) = \bar{f}(A)\bar{g}(V) + \bar{f}(V)\bar{g}(A) - \bar{f}(A)\bar{g}(A)$.

Theorem (Guillory & Bilmes, 2011)

$h(A)$ so defined is polymatroidal.

The "or" of two polymatroid functions

- Given two polymatroid functions f and g, suppose feasible A are defined as $\{A : f(A) \geq \alpha_f \text{ or } g(A) \geq \alpha_g\}$ for real α_f, α_g.
- This is identical to: $\left\{A : \bar{f}(A) = \alpha_f \text{ or } \bar{g}(A) = \alpha_g\right\}$ where $\bar{f}(A) = \min(f(A), \alpha_f)$ and $\bar{g}(A) = \min(f(A), \alpha_g)$
- Define: $h(A) = \bar{f}(A)\bar{g}(V) + \bar{f}(V)\bar{g}(A) - \bar{f}(A)\bar{g}(A)$.

Theorem (Guillory & Bilmes, 2011)

$h(A)$ so defined is polymatroidal.

Theorem

$h(A) = \alpha_f\alpha_g$ if and only if $\bar{f}(A) = \alpha_f$ or $\bar{g}(A) = \alpha_g$

The "or" of two polymatroid functions

- Given two polymatroid functions f and g, suppose feasible A are defined as $\{A : f(A) \geq \alpha_f \text{ or } g(A) \geq \alpha_g\}$ for real α_f, α_g.
- This is identical to: $\{A : \bar{f}(A) = \alpha_f \text{ or } \bar{g}(A) = \alpha_g\}$ where $\bar{f}(A) = \min(f(A), \alpha_f)$ and $\bar{g}(A) = \min(f(A), \alpha_g)$
- Define: $h(A) = \bar{f}(A)\bar{g}(V) + \bar{f}(V)\bar{g}(A) - \bar{f}(A)\bar{g}(A)$.

Theorem (Guillory & Bilmes, 2011)

$h(A)$ so defined is polymatroidal.

Theorem

$h(A) = \alpha_f\alpha_g$ *if and only if* $\bar{f}(A) = \alpha_f$ *or* $\bar{g}(A) = \alpha_g$

- Therefore, h can be used as a submodular surrogate for the "or" of multiple submodular functions.

Composition and Submodular Functions

- Convex/Concave have many nice properties of composition (see Boyd & Vandenberghe, "Convex Optimization")

Composition and Submodular Functions

- Convex/Concave have many nice properties of composition (see Boyd & Vandenberghe, "Convex Optimization")
- A submodular function $f : 2^V \to \mathbb{R}$ has a different type of input and output, so composing two submodular functions directly makes no sense.

Composition and Submodular Functions

- Convex/Concave have many nice properties of composition (see Boyd & Vandenberghe, "Convex Optimization")
- A submodular function $f : 2^V \to \mathbb{R}$ has a different type of input and output, so composing two submodular functions directly makes no sense.
- However, we have a number of forms of composition results that preserve submodularity, which we turn to next:

Grouping elements, set cover, and bipartite neighborhoods

- Given submodular $f : 2^V \to \mathbb{R}$ and a grouping of $V = V_1 \cup V_2 \cup \cdots \cup V_k$ into k possibly overlapping clusters.

Grouping elements, set cover, and bipartite neighborhoods

- Given submodular $f : 2^V \to \mathbb{R}$ and a grouping of $V = V_1 \cup V_2 \cup \cdots \cup V_k$ into k possibly overlapping clusters.
- Define new function $g : 2^{[k]} \to \mathbb{R}$ where $\forall D \subseteq [k] = \{1, 2, \ldots, k\}$,

$$g(D) = f(\bigcup_{d \in D} V_d) \tag{41}$$

Grouping elements, set cover, and bipartite neighborhoods

- Given submodular $f : 2^V \to \mathbb{R}$ and a grouping of $V = V_1 \cup V_2 \cup \dots \cup V_k$ into k possibly overlapping clusters.
- Define new function $g : 2^{[k]} \to \mathbb{R}$ where $\forall D \subseteq [k] = \{1, 2, \dots, k\}$,

$$g(D) = f(\bigcup_{d \in D} V_d) \tag{41}$$

- Then g is submodular if either f is monotone non-decreasing or the sets $\{V_i\}$ are disjoint.

Grouping elements, set cover, and bipartite neighborhoods

- Given submodular $f : 2^V \to \mathbb{R}$ and a grouping of $V = V_1 \cup V_2 \cup \dots \cup V_k$ into k possibly overlapping clusters.
- Define new function $g : 2^{[k]} \to \mathbb{R}$ where $\forall D \subseteq [k] = \{1, 2, \dots, k\}$,

$$g(D) = f(\bigcup_{d \in D} V_d) \tag{41}$$

- Then g is submodular if either f is monotone non-decreasing or the sets $\{V_i\}$ are disjoint.
- Ex: Bipartite neighborhoods: Let $\Gamma : 2^V \to \mathbb{R}$ be the neighbor function in a bipartite graph $G = (V, U, E, w)$. V is set of "left" nodes, U is set of right nodes, $E \subseteq V \times U$ are edges, and $w : 2^E \to \mathbb{R}$ is a modular function on edges.

Grouping elements, set cover, and bipartite neighborhoods

- Given submodular $f : 2^V \to \mathbb{R}$ and a grouping of $V = V_1 \cup V_2 \cup \cdots \cup V_k$ into k possibly overlapping clusters.
- Define new function $g : 2^{[k]} \to \mathbb{R}$ where $\forall D \subseteq [k] = \{1, 2, \ldots, k\}$,

$$g(D) = f(\bigcup_{d \in D} V_d) \tag{41}$$

- Then g is submodular if either f is monotone non-decreasing or the sets $\{V_i\}$ are disjoint.
- Ex: Bipartite neighborhoods: Let $\Gamma : 2^V \to \mathbb{R}$ be the neighbor function in a bipartite graph $G = (V, U, E, w)$. V is set of "left" nodes, U is set of right nodes, $E \subseteq V \times U$ are edges, and $w : 2^E \to \mathbb{R}$ is a modular function on edges.
- Neighbors defined as $\Gamma(X) = \{u \in U : |X \times \{u\} \cap E| \geq 1\}$ for $X \subseteq V$.

Grouping elements, set cover, and bipartite neighborhoods

- Given submodular $f : 2^V \to \mathbb{R}$ and a grouping of $V = V_1 \cup V_2 \cup \dots \cup V_k$ into k possibly overlapping clusters.
- Define new function $g : 2^{[k]} \to \mathbb{R}$ where $\forall D \subseteq [k] = \{1, 2, \dots, k\}$,

$$g(D) = f\Big(\bigcup_{d \in D} V_d\Big) \tag{41}$$

- Then g is submodular if either f is monotone non-decreasing or the sets $\{V_i\}$ are disjoint.
- Ex: Bipartite neighborhoods: Let $\Gamma : 2^V \to \mathbb{R}$ be the neighbor function in a bipartite graph $G = (V, U, E, w)$. V is set of "left" nodes, U is set of right nodes, $E \subseteq V \times U$ are edges, and $w : 2^E \to \mathbb{R}$ is a modular function on edges.
- Neighbors defined as $\Gamma(X) = \{u \in U : |X \times \{u\} \cap E| \geq 1\}$ for $X \subseteq V$. Then $f(\Gamma(X))$ is submodular. Special case: set cover.

Grouping elements, set cover, and bipartite neighborhoods

- Given submodular $f : 2^V \to \mathbb{R}$ and a grouping of $V = V_1 \cup V_2 \cup \cdots \cup V_k$ into k possibly overlapping clusters.
- Define new function $g : 2^{[k]} \to \mathbb{R}$ where $\forall D \subseteq [k] = \{1, 2, \ldots, k\}$,

$$g(D) = f\Big(\bigcup_{d \in D} V_d\Big) \tag{41}$$

- Then g is submodular if either f is monotone non-decreasing or the sets $\{V_i\}$ are disjoint.
- Ex: Bipartite neighborhoods: Let $\Gamma : 2^V \to \mathbb{R}$ be the neighbor function in a bipartite graph $G = (V, U, E, w)$. V is set of "left" nodes, U is set of right nodes, $E \subseteq V \times U$ are edges, and $w : 2^E \to \mathbb{R}$ is a modular function on edges.
- Neighbors defined as $\Gamma(X) = \{u \in U : |X \times \{u\} \cap E| \geq 1\}$ for $X \subseteq V$. Then $f(\Gamma(X))$ is submodular. Special case: set cover.
- In fact, all integral polymatroid functions can be obtained in g above for f a matroid rank function and $\{V_d\}$ appropriately chosen.

Concave composed with polymatroid

We also have the following composition property with concave functions:

Theorem

Given functions $f : 2^V \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$, the composition $h = f \circ g : 2^V \to \mathbb{R}$ (i.e., $h(S) = g(f(S))$) is nondecreasing submodular, if g is non-decreasing concave and f is nondecreasing submodular.

Concave composed with non-negative modular

Theorem

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

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

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

Concave composed with non-negative modular

Theorem

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

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

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

$$f(A) = \sum_{i=1}^{K} g_i(m_i(A)) \tag{42}$$

Concave composed with non-negative modular

Theorem

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

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

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

$$f(A) = \sum_{i=1}^{K} g_i(m_i(A)) \tag{42}$$

- Very large class of functions, including graph cut, bipartite neighborhoods, set cover (Stobbe & Krause).

Concave composed with non-negative modular

Theorem

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

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

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

$$f(A) = \sum_{i=1}^{K} g_i(m_i(A)) \tag{42}$$

- Very large class of functions, including graph cut, bipartite neighborhoods, set cover (Stobbe & Krause).
- However, Vondrak showed that a graphic matroid rank function over K_4 can't be represented in this fashion.

Weighted Matroid Rank Functions

- We saw matroid rank is submodular. Given matroid $(V, \mathcal{I})$,

$$f(B) = \max \{|A| : A \subseteq B \text{ and } A \in \mathcal{I}\} \tag{43}$$

Weighted Matroid Rank Functions

- We saw matroid rank is submodular. Given matroid $(V, \mathcal{I})$,

$$f(B) = \max \{|A| : A \subseteq B \text{ and } A \in \mathcal{I}\} \tag{43}$$

- Weighted matroid rank functions. Given matroid $(V, \mathcal{I})$, and non-negative modular function $m : 2^V \to \mathbb{R}_+$,

$$f(B) = \max \{m(A) : A \subseteq B \text{ and } A \in \mathcal{I}\} \tag{44}$$

  is also submodular.

Weighted Matroid Rank Functions

- We saw matroid rank is submodular. Given matroid $(V, \mathcal{I})$,

$$f(B) = \max \{|A| : A \subseteq B \text{ and } A \in \mathcal{I}\} \tag{43}$$

- Weighted matroid rank functions. Given matroid $(V, \mathcal{I})$, and non-negative modular function $m : 2^V \to \mathbb{R}_+$,

$$f(B) = \max \{m(A) : A \subseteq B \text{ and } A \in \mathcal{I}\} \tag{44}$$

  is also submodular.
- Take a 1-partition matroid with limit k, we get:

$$f(B) = \max \{m(A) : A \subseteq B \text{ and } |A| \leq k\} \tag{45}$$

Weighted Matroid Rank Functions

- We saw matroid rank is submodular. Given matroid $(V, \mathcal{I})$,

$$f(B) = \max\{|A| : A \subseteq B \text{ and } A \in \mathcal{I}\} \tag{43}$$

- Weighted matroid rank functions. Given matroid $(V, \mathcal{I})$, and non-negative modular function $m : 2^V \to \mathbb{R}_+$,

$$f(B) = \max\{m(A) : A \subseteq B \text{ and } A \in \mathcal{I}\} \tag{44}$$

  is also submodular.

- Take a 1-partition matroid with limit k, we get:

$$f(B) = \max\{m(A) : A \subseteq B \text{ and } |A| \leq k\} \tag{45}$$

- Take a 1-partition matroid with limit 1, we get the max function:

$$f(B) = \max_{b \in B} m(b) \tag{46}$$

Facility Location

- Given a set of k matroids $(V, \mathcal{I}_i)$ and k modular weight functions m_i, the following is submodular:

$$f(A) = \sum_{i=1}^{k} \alpha_i \max\{m_i(A) : A \subseteq B \text{ and } A \in \mathcal{I}_i\} \tag{47}$$

Facility Location

- Given a set of k matroids $(V, \mathcal{I}_i)$ and k modular weight functions m_i, the following is submodular:

$$f(A) = \sum_{i=1}^{k} \alpha_i \max\{m_i(A) : A \subseteq B \text{ and } A \in \mathcal{I}_i\} \tag{47}$$

- Take all $\alpha_i = 1$, all matroids 1-partition matroids, and set $w_{ij} = m_i(j)$, and $k = |V|$ for some weighted graph $G = (V, E, w)$, we get the uncapacitated facility location function:

$$f(A) = \sum_{i \in V} \max_{a \in A} w_{ai} \tag{48}$$

Information and Complexity functions

- Given a set V of items, we might wish to measure the "information" or "complexity" in a subset $A \subset V$.

Information and Complexity functions

- Given a set V of items, we might wish to measure the "information" or "complexity" in a subset $A \subset V$.
- Matroid rank $r(A)$ can measure the "information" or "complexity" via the dimensionality spanned by vectors with indices A.

Information and Complexity functions

- Given a set V of items, we might wish to measure the "information" or "complexity" in a subset $A \subset V$.
- Matroid rank $r(A)$ can measure the "information" or "complexity" via the dimensionality spanned by vectors with indices A.
- Unit increment $r(v|A) \in \{0, 1\}$ so no partial independence.

Information and Complexity functions

- Given a set V of items, we might wish to measure the "information" or "complexity" in a subset $A \subset V$.
- Matroid rank $r(A)$ can measure the "information" or "complexity" via the dimensionality spanned by vectors with indices A.
- Unit increment $r(v|A) \in \{0, 1\}$ so no partial independence.
- Entropy of a set of random variables $\{X_v\}_{v \in V}$, where

$$f(A) = H(X_A) = H(\bigcup_{a \in A} X_a) = -\sum_{x_A} \Pr(x_A) \log \Pr(x_A) \tag{49}$$

can measure partial independence.

Information and Complexity functions

- Given a set V of items, we might wish to measure the "information" or "complexity" in a subset $A \subset V$.
- Matroid rank $r(A)$ can measure the "information" or "complexity" via the dimensionality spanned by vectors with indices A.
- Unit increment $r(v|A) \in \{0, 1\}$ so no partial independence.
- Entropy of a set of random variables $\{X_v\}_{v \in V}$, where

$$f(A) = H(X_A) = H(\bigcup_{a \in A} X_a) = -\sum_{x_A} \Pr(x_A) \log \Pr(x_A) \tag{49}$$

  can measure partial independence.
- Entropy is submodular due to non-negativity of conditional mutual information. Given $A, B, C \subseteq V$,

$$\begin{aligned} &I(X_{A\setminus B}; X_{B\setminus A} | X_{A\cap B}) \\ &\quad = H(X_A) + H(X_B) - H(X_{A\cup B}) - H(X_{A\cap B}) \geq 0 \end{aligned} \tag{50}$$

Generalized information/complexity functions

- Entropy requires a joint probability distribution over items, while rank requires a vector space.

Generalized information/complexity functions

- Entropy requires a joint probability distribution over items, while rank requires a vector space.
- Many information functions are statistical, requiring a distribution, and measure information within a distribution. E.g., entropy, Rényi's information, Daroczy's entropy, etc.

Generalized information/complexity functions

- Entropy requires a joint probability distribution over items, while rank requires a vector space.
- Many information functions are statistical, requiring a distribution, and measure information within a distribution. E.g., entropy, Rényi's information, Daroczy's entropy, etc.
- Some require a generating algorithm (Kolmogorov complexity).

Generalized information/complexity functions

- Entropy requires a joint probability distribution over items, while rank requires a vector space.
- Many information functions are statistical, requiring a distribution, and measure information within a distribution. E.g., entropy, Rényi's information, Daroczy's entropy, etc.
- Some require a generating algorithm (Kolmogorov complexity).
- Submodularity is a natural property of an "information" or "complexity" function over subsets of objects.

Generalized information/complexity functions

- Entropy requires a joint probability distribution over items, while rank requires a vector space.
- Many information functions are statistical, requiring a distribution, and measure information within a distribution. E.g., entropy, Rényi's information, Daroczy's entropy, etc.
- Some require a generating algorithm (Kolmogorov complexity).
- Submodularity is a natural property of an "information" or "complexity" function over subsets of objects.
- All submodular functions express a form of "abstract independence" or "generalized complexity"

Polymatroids: Generalized Dependence

- there is a notion of "independence" , i.e., $A \perp\!\!\!\perp B$:

$$f(A \cup B) = f(A) + f(B), \tag{51}$$

Polymatroids: Generalized Dependence

- there is a notion of "independence" , i.e., $A \perp\!\!\!\perp B$:

$$f(A \cup B) = f(A) + f(B), \tag{51}$$

- and a notion of "conditional independence" , i.e., $A \perp\!\!\!\perp B | C$:

$$f(A \cup B \cup C) + f(C) = f(A \cup C) + f(B \cup C) \tag{52}$$

Polymatroids: Generalized Dependence

- there is a notion of “independence” , i.e., $A \perp\!\!\!\perp B$:

$$f(A \cup B) = f(A) + f(B), \tag{51}$$

- and a notion of “conditional independence” , i.e., $A \perp\!\!\!\perp B | C$:

$$f(A \cup B \cup C) + f(C) = f(A \cup C) + f(B \cup C) \tag{52}$$

- and a notion of “dependence” (conditioning reduces valuation):

$$f(A|B) \triangleq f(A \cup B) - f(B) < f(A), \tag{53}$$

Polymatroids: Generalized Dependence

- there is a notion of "independence" , i.e., $A \perp\!\!\!\perp B$:

$$f(A \cup B) = f(A) + f(B), \tag{51}$$

- and a notion of "conditional independence" , i.e., $A \perp\!\!\!\perp B | C$:

$$f(A \cup B \cup C) + f(C) = f(A \cup C) + f(B \cup C) \tag{52}$$

- and a notion of "dependence" (conditioning reduces valuation):

$$f(A|B) \triangleq f(A \cup B) - f(B) < f(A), \tag{53}$$

- and a notion of "conditional mutual information"

$$I_f(A; B|C) \triangleq f(A \cup C) + f(B \cup C) - f(A \cup B \cup C) - f(C) \geq 0$$

Polymatroids: Generalized Dependence

- there is a notion of "independence" , i.e., $A \perp\!\!\!\perp B$:

$$f(A \cup B) = f(A) + f(B), \tag{51}$$

- and a notion of "conditional independence" , i.e., $A \perp\!\!\!\perp B | C$:

$$f(A \cup B \cup C) + f(C) = f(A \cup C) + f(B \cup C) \tag{52}$$

- and a notion of "dependence" (conditioning reduces valuation):

$$f(A|B) \triangleq f(A \cup B) - f(B) < f(A), \tag{53}$$

- and a notion of "conditional mutual information"

$$I_f(A; B|C) \triangleq f(A \cup C) + f(B \cup C) - f(A \cup B \cup C) - f(C) \geq 0$$

- and two notions of "information amongst a collection of sets":

$$I_f(S_1; S_2; \ldots; S_k) = \sum_{i=1}^{k} f(S_k) - f(S_1 \cup S_2 \cup \cdots \cup S_k) \tag{54}$$

$$I'_f(S_1; S_2; \ldots; S_k) = \sum_{A \subseteq \{1,2,\ldots,k\}} (-1)^{|A|+1} f(\bigcup_{j \in A} S_j) \tag{55}$$

Submodular Separation and Symmetric Submodular Minimization

- Subsets A and B are separable if $f(A \cup B) = f(A) + f(B)$
- Hence, separability is the same as statistical independence when f is the entropy function.
- Partitioning V into separable blocks can be performed using symmetric SFM.
- Given any polymatroid f, symmetrize it as follows:

Symmetric Submodular Functions

- Symmetrize and normalize f as $f \to \check{f}$ via the operation:

$$\check{f}(A) = f(A) + f(V \setminus A) - f(V), \tag{56}$$

so that $\check{f}(\emptyset) = 0$ if $f(\emptyset) = 0$, and $\check{f}(A) = \check{f}(V \setminus A)$ for all A.

Symmetric Submodular Functions

- Symmetrize and normalize f as $f \to \breve{f}$ via the operation:

$$\breve{f}(A) = f(A) + f(V \setminus A) - f(V), \tag{56}$$

so that $\breve{f}(\emptyset) = 0$ if $f(\emptyset) = 0$, and $\breve{f}(A) = \breve{f}(V \setminus A)$ for all A.

- Such an $\breve{f}$ is also non-negative since

$$2\breve{f}(A) = \breve{f}(A) + \breve{f}(V \setminus A) \geq \breve{f}(\emptyset) + \breve{f}(V) = 2\breve{f}(\emptyset) \geq 0 \tag{57}$$

Symmetric Submodular Functions

- Symmetrize and normalize f as $f \to \check{f}$ via the operation:

$$\check{f}(A) = f(A) + f(V \setminus A) - f(V), \tag{56}$$

so that $\check{f}(\emptyset) = 0$ if $f(\emptyset) = 0$, and $\check{f}(A) = \check{f}(V \setminus A)$ for all A.

- Such an $\check{f}$ is also non-negative since

$$2\check{f}(A) = \check{f}(A) + \check{f}(V \setminus A) \geq \check{f}(\emptyset) + \check{f}(V) = 2\check{f}(\emptyset) \geq 0 \tag{57}$$

- Any submodular function can be so symmetrized, and submodularity is preserved.

Symmetric Submodular Functions

- Symmetrize and normalize f as $f \to \check{f}$ via the operation:

$$\check{f}(A) = f(A) + f(V \setminus A) - f(V), \tag{56}$$

so that $\check{f}(\emptyset) = 0$ if $f(\emptyset) = 0$, and $\check{f}(A) = \check{f}(V \setminus A)$ for all A.

- Such an $\check{f}$ is also non-negative since

$$2\check{f}(A) = \check{f}(A) + \check{f}(V \setminus A) \geq \check{f}(\emptyset) + \check{f}(V) = 2\check{f}(\emptyset) \geq 0 \tag{57}$$

- Any submodular function can be so symmetrized, and submodularity is preserved.
- Example: $f(A) = H(X_A) =$ entropy, then $\check{f}(A) = I(X_A; X_{V \setminus A}) =$ symmetric mutual information.

Separators of submodular function via symmetrized version

- Such a symmetrized submodular function measures a form of "dependence" between A and $\bar{A} \triangleq V \setminus A$

Separators of submodular function via symmetrized version

- Such a symmetrized submodular function measures a form of "dependence" between A and $\bar{A} \triangleq V \setminus A$

Theorem

We are given an f that is normalized & submodular. If $\exists A$ such that:

$$\check{f}(A) \triangleq f(A) + f(\bar{A}) - f(V) = 0 \tag{58}$$

then f is "decomposable" w.r.t. A. This means that $f(B) = f(B \cap A) + f(B \cap \bar{A})$ for all B.

Gaussian entropy, and the log-determinant function

Definition (differential entropy $h(X)$)

$$h(X) = -\int_S f(x) \log f(x) dx \tag{59}$$

- When $x \sim \mathcal{N}(\mu, \Sigma)$ is multivariate Gaussian, the (differential) entropy of the r.v. X is given by

$$h(X) = \log \sqrt{|2\pi e \mathbf{\Sigma}|} = \log \sqrt{(2\pi e)^n |\mathbf{\Sigma}|} \tag{60}$$

and in particular, for a variable subset A and a constant γ,

$$f(A) = h(X_A) = \log \sqrt{(2\pi e)^{|A|} |\mathbf{\Sigma}_A|} = \gamma |A| + \frac{1}{2} \log |\mathbf{\Sigma}_A| \tag{61}$$

Gaussian entropy, and the log-determinant function

Definition (differential entropy $h(X)$)

$$h(X) = -\int_S f(x) \log f(x) dx \tag{59}$$

- When $x \sim \mathcal{N}(\mu, \Sigma)$ is multivariate Gaussian, the (differential) entropy of the r.v. X is given by

$$h(X) = \log \sqrt{|2\pi e \mathbf{\Sigma}|} = \log \sqrt{(2\pi e)^n |\mathbf{\Sigma}|} \tag{60}$$

and in particular, for a variable subset A and a constant γ,

$$f(A) = h(X_A) = \log \sqrt{(2\pi e)^{|A|} |\mathbf{\Sigma}_A|} = \gamma |A| + \frac{1}{2} \log |\mathbf{\Sigma}_A| \tag{61}$$

- Application of Jensen's inequality shows that $I(X_{A\setminus B}; X_{B\setminus A} | X_{A\cap B}) = h(X_A) + h(X_B) - h(X_{A\cup B}) - h(X_{A\cap B}) \geq 0$. Hence differential entropy is submodular, and thus so is the logdet function.

Are all polymatroid functions entropy functions?

- No, entropy functions must also satisfy the following:

Are all polymatroid functions entropy functions?

- No, entropy functions must also satisfy the following:

Theorem (Yeung)

For any four discrete random variables $\{X, Y, Z, U\}$*, then*

$$I(X;Y) = I(X;Y|Z) = 0 \tag{62}$$

implies that

$$I(X;Y|Z,U) \le I(Z;U|X,Y) + I(X;Y|U) \tag{63}$$

where $I(\cdot;\cdot|\cdot)$ *is the standard Shannon mutual information function.*

Are all polymatroid functions entropy functions?

- No, entropy functions must also satisfy the following:

Theorem (Yeung)

For any four discrete random variables $\{X, Y, Z, U\}$, then

$$I(X; Y) = I(X; Y|Z) = 0 \tag{62}$$

implies that

$$I(X; Y|Z, U) \leq I(Z; U|X, Y) + I(X; Y|U) \tag{63}$$

where $I(\cdot; \cdot|\cdot)$ is the standard Shannon mutual information function.

- This is not required for all polymatroid-based conditional mutual information functions $I_f(\cdot; \cdot|\cdot)$.

Containment, Gaussian Entropy, and DPPs

- Submodular functions $\supset$ Polymatroid functions $\supset$ Entropy functions $\supset$ Gaussian Entropy functions $=$ DPPs.

Containment, Gaussian Entropy, and DPPs

- Submodular functions $\supset$ Polymatroid functions $\supset$ Entropy functions $\supset$ Gaussian Entropy functions $=$ DPPs.
- DPP is a point process where $\Pr(\mathbf{Y} = Y) \propto \det(L_Y)$ for some positive-definite matrix L, so DPPs are log-submodular.

Containment, Gaussian Entropy, and DPPs

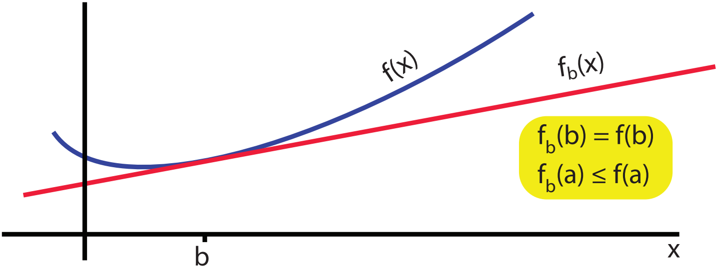
- Submodular functions $\supset$ Polymatroid functions $\supset$ Entropy functions $\supset$ Gaussian Entropy functions $=$ DPPs.
- DPP is a point process where $\Pr(\mathbf{Y} = Y) \propto \det(L_Y)$ for some positive-definite matrix L, so DPPs are log-submodular.
- Thanks to the properties of matrix algebra (e.g., determinants), DPPs are computationally extremely attractive.

Containment, Gaussian Entropy, and DPPs

- Submodular functions $\supset$ Polymatroid functions $\supset$ Entropy functions $\supset$ Gaussian Entropy functions $=$ DPPs.
- DPP is a point process where $\Pr(\mathbf{Y} = Y) \propto \det(L_Y)$ for some positive-definite matrix L, so DPPs are log-submodular.
- Thanks to the properties of matrix algebra (e.g., determinants), DPPs are computationally extremely attractive.
- We'll touch DPPs a bit later when we compare submodularity and graphical models.

Outline

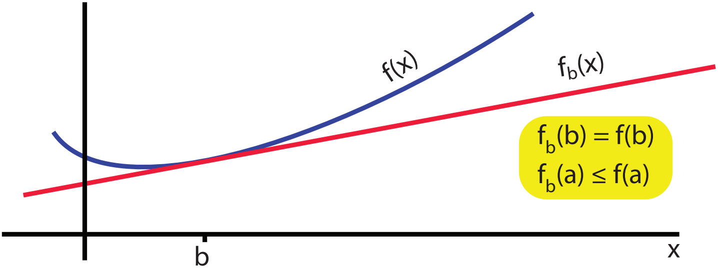
Convex Functions and Tight Subgradients



- A convex function f has a subgradient at any in-domain point b, namely there exists f_b such that

$$f(x) - f(b) \geq \langle f_b, x - b \rangle, \forall x. \tag{64}$$

Concave Functions and Tight Supergradients

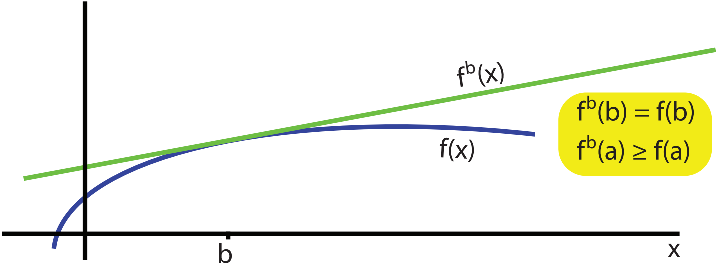


- A convex function f has a subgradient at any in-domain point b, namely there exists f_b such that

$$f(x) - f(b) \geq \langle f_b, x - b \rangle, \forall x. \tag{64}$$

- We have that $f(x)$ is convex, $f_b(x)$ is affine, and is a tight subgradient (tight at b, affine lower bound on $f(x)$).

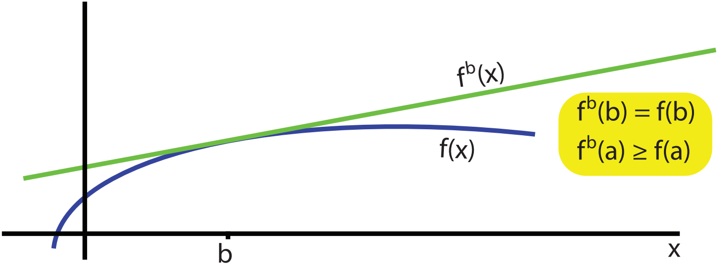
Convex Functions and Tight Subgradients



- A concave f has a supergradient at any in-domain point b, namely there exists f^b such that

$$f(x) - f(b) \leq \langle f^b, x - b \rangle, \forall x. \tag{65}$$

Concave Functions and Tight Supergradients



- A concave f has a supergradient at any in-domain point b, namely there exists f^b such that

$$f(x) - f(b) \leq \langle f^b, x - b \rangle, \forall x. \tag{65}$$

- We have that $f(x)$ is concave, $f^b(x)$ is affine, and is a tight supergradient (tight at b, affine upper bound on $f(x)$).

Trivial additive upper/lower bounds

- Any submodular function has trivial additive upper and lower bounds. That is for all $A \subseteq V$,

$$m_f(A) \leq f(A) \leq m^f(A) \tag{66}$$

where

$$m^f(A) = \sum_{a \in A} f(a) \tag{67}$$

$$m_f(A) = \sum_{a \in A} f(a|V \setminus \{a\}) \tag{68}$$

Trivial additive upper/lower bounds

- Any submodular function has trivial additive upper and lower bounds. That is for all $A \subseteq V$,

$$m_f(A) \leq f(A) \leq m^f(A) \tag{66}$$

where

$$m^f(A) = \sum_{a \in A} f(a) \tag{67}$$

$$m_f(A) = \sum_{a \in A} f(a | V \setminus \{a\}) \tag{68}$$

- $m^f \in \mathbb{R}^V$ and $m_f \in \mathbb{R}^V$ are both modular (or additive) functions.

Trivial additive upper/lower bounds

- Any submodular function has trivial additive upper and lower bounds. That is for all $A \subseteq V$,

$$m_f(A) \leq f(A) \leq m^f(A) \tag{66}$$

where

$$m^f(A) = \sum_{a \in A} f(a) \tag{67}$$

$$m_f(A) = \sum_{a \in A} f(a|V \setminus \{a\}) \tag{68}$$

- $m^f \in \mathbb{R}^V$ and $m_f \in \mathbb{R}^V$ are both modular (or additive) functions.
- A "semigradient" is customized, and at least at one point is tight.

Submodular Subgradients

- For submodular function f, the subdifferential (all subgradients tight at $X \subseteq V$) can be defined as:

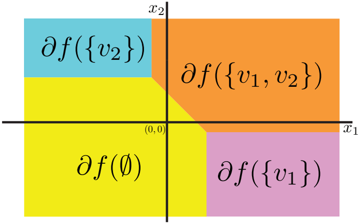
$$\partial f(X) = \{x \in \mathbb{R}^V : \forall Y \subseteq V, x(Y) - x(X) \leq f(Y) - f(X)\} \quad (69)$$

Submodular Subgradients

- For submodular function f, the subdifferential (all subgradients tight at $X \subseteq V$) can be defined as:

$$\partial f(X) = \{x \in \mathbb{R}^V : \forall Y \subseteq V, x(Y) - x(X) \leq f(Y) - f(X)\} \quad (69)$$

- This partitions $\mathbb{R}^V$:

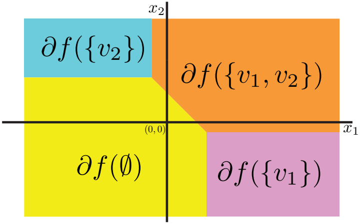


Submodular Subgradients

- For submodular function f, the subdifferential (all subgradients tight at $X \subseteq V$) can be defined as:

$$\partial f(X) = \{x \in \mathbb{R}^V : \forall Y \subseteq V, x(Y) - x(X) \leq f(Y) - f(X)\} \quad (69)$$

- This partitions $\mathbb{R}^V$:



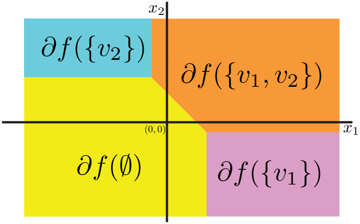
- Extreme points are easy to get via Edmonds's greedy algorithm:

Submodular Subgradients

- For submodular function f, the subdifferential (all subgradients tight at $X \subseteq V$) can be defined as:

$$\partial f(X) = \{x \in \mathbb{R}^V : \forall Y \subseteq V, x(Y) - x(X) \leq f(Y) - f(X)\} \quad (69)$$

- This partitions $\mathbb{R}^V$:



- Extreme points are easy to get via Edmonds's greedy algorithm:

Theorem (Fujishige 2005, Theorem 6.11)

A point $y \in \mathbb{R}^V$ is an extreme point of $\partial f(X)$, iff there exists a maximal chain $\emptyset = S_0 \subset S_1 \subset \cdots \subset S_n$ with $X = S_j$ for some j, such that $y(S_i \setminus S_{i-1}) = y(S_i) - y(S_{i-1}) = f(S_i) - f(S_{i-1})$.

The Submodular Subgradients (Fujishige 2005)

- For an arbitrary $Y \subseteq V$
- Let σ be a permutation of V and define $S_i^\sigma = \{\sigma(1), \sigma(2), \dots, \sigma(i)\}$ as σ's chain where $S_k^\sigma = Y$ where $|Y| = k$.
- We can define a subgradient h_Y^f corresponding to f as:

$$h_{Y,\sigma}^f(\sigma(i)) = \begin{cases} f(S_1^\sigma) & \text{if } i = 1 \\ f(S_i^\sigma) - f(S_{i-1}^\sigma) & \text{otherwise} \end{cases} .$$

- We get a tight modular lower bound of f as follows:

$$h_{Y,\sigma}^f(X) \triangleq \sum_{x \in X} h_{Y,\sigma}^f(x) \leq f(X), \forall X \subseteq V.$$

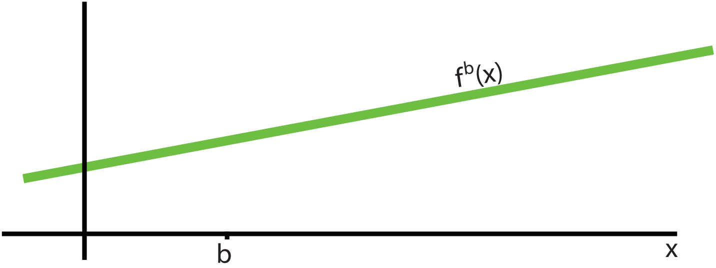
Note, tight at Y means $h_{Y,\sigma}^f(Y) = f(Y)$.

Convexity and Tight Sub- and Super-gradients?

- Can there be both a tight linear upper bound and tight linear lower bound on a convex (or concave) function, where each bound is tight at the same point?

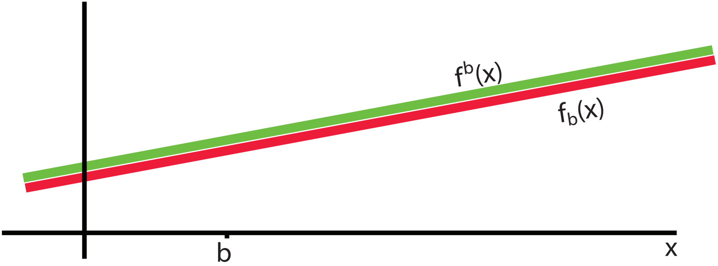
Convexity and Tight Sub- and Super-gradients?

- Can there be both a tight linear upper bound and tight linear lower bound on a convex (or concave) function, where each bound is tight at the same point?



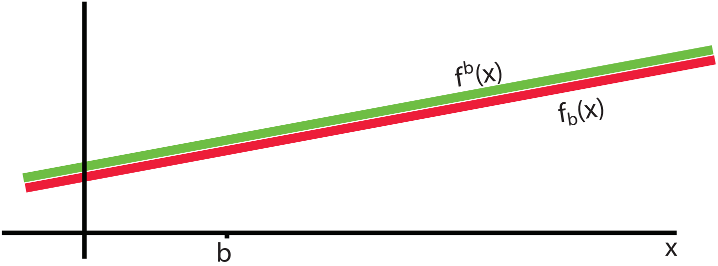
Convexity and Tight Sub- and Super-gradients?

- Can there be both a tight linear upper bound and tight linear lower bound on a convex (or concave) function, where each bound is tight at the same point?



Convexity and Tight Sub- and Super-gradients?

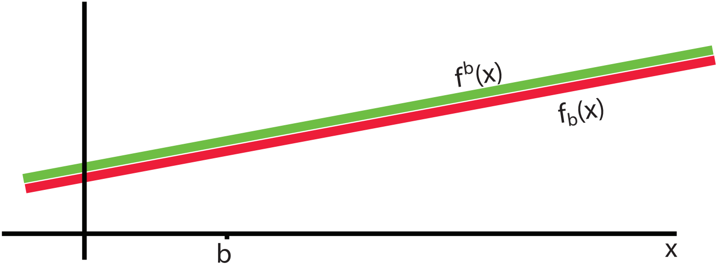
- Can there be both a tight linear upper bound and tight linear lower bound on a convex (or concave) function, where each bound is tight at the same point?



- If a continuous function has both a sub- and super-gradient at a point, then the function must be affine.

Convexity and Tight Sub- and Super-gradients?

- Can there be both a tight linear upper bound and tight linear lower bound on a convex (or concave) function, where each bound is tight at the same point?



- If a continuous function has both a sub- and super-gradient at a point, then the function must be affine.
- What about discrete set functions?

The Submodular Supergradients

- Can a submodular function also have a supergradient? We saw that in the continuous case, simultaneous sub/super gradients meant linear.
- (Nemhauser, Wolsey, & Fisher 1978) established the following iff conditions for submodularity (if either hold, f is submodular):

$$f(Y) \le f(X) - \sum_{j \in X \setminus Y} f(j|X \setminus j) + \sum_{j \in Y \setminus X} f(j|X \cap Y),$$

$$f(Y) \le f(X) - \sum_{j \in X \setminus Y} f(j|(X \cup Y) \setminus j) + \sum_{j \in Y \setminus X} f(j|X)$$

Recall that $f(A|B) \triangleq f(A \cup B) - f(B)$ is the gain of adding A in the context of B.

Submodular and Supergradients

- Using submodularity further, these can be relaxed to produce two tight modular upper bounds (Jegelka & Bilmes, 2011, Iyer & Bilmes 2013):

$$f(Y) \leq m^f_{X,1}(Y) \triangleq f(X) - \sum_{j \in X \setminus Y} f(j|X \setminus j) + \sum_{j \in Y \setminus X} f(j|\emptyset),$$

$$f(Y) \leq m^f_{X,2}(Y) \triangleq f(X) - \sum_{j \in X \setminus Y} f(j|V \setminus j) + \sum_{j \in Y \setminus X} f(j|X).$$

Hence, this yields three tight (at set X) modular upper bounds $m^f_{X,1}, m^f_{X,2}$ for any submodular function f.

Why is $m^f_{X,2}$ Modular?

- $m : 2^V \rightarrow \mathbb{R}$ is modular if $m(X) + m(Y) = m(X \cup Y) + m(X \cap Y)$, or equivalently if it can be expressed as, for any $X \subseteq V$:

$$m(X) = c + \sum_{j \in X} m(j) \tag{70}$$

where c is a constant. I.e., $m \in \mathbb{R}^V$.

- For example, the function

$$m^f_{X,2}(Y) \triangleq f(X) - \sum_{j \in X \setminus Y} f(j|V \setminus j) + \sum_{j \in Y \setminus X} f(j|X) \tag{71}$$

is modular in Y as Equation (70) with

$$m^f_{X,2}(Y) \triangleq \left[f(X) - \sum_{j \in X} f(j|V \setminus j) \right] + \sum_{j \in (X \cap Y)} f(j|V \setminus j) \tag{72}$$

$$+ \sum_{j \in Y \setminus X} f(j|X) \tag{73}$$

Submodular Supergradients (Iyer et al, 2013)

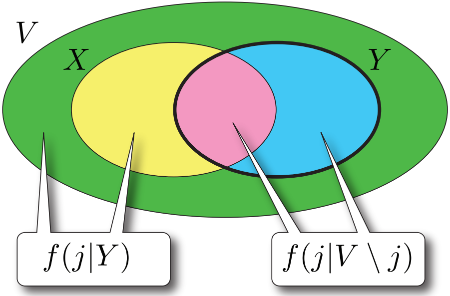
- These three supergradients (which we call grow, shrink, and bar) are easy to obtain.

Submodular Supergradients (Iyer et al, 2013)

- These three supergradients (which we call grow, shrink, and bar) are easy to obtain.

Grow:

$$\hat{g}_Y(j) = \begin{cases} f(j|Y) & \text{for } j \notin Y \\ f(j|V \setminus \{j\}) & \text{for } j \in Y \end{cases}$$

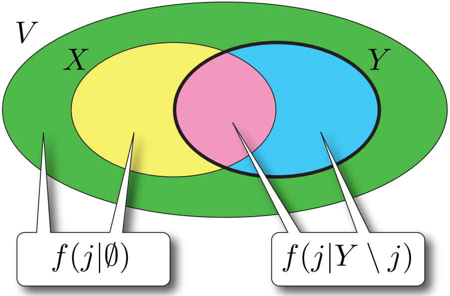


Submodular Supergradients (Iyer et al, 2013)

- These three supergradients (which we call grow, shrink, and bar) are easy to obtain.

Shrink:

$$\check{g}_Y(j) = \begin{cases} f(j|\emptyset) & \text{for } j \notin Y \\ f(j|Y \setminus \{j\}) & \text{for } j \in Y \end{cases}$$

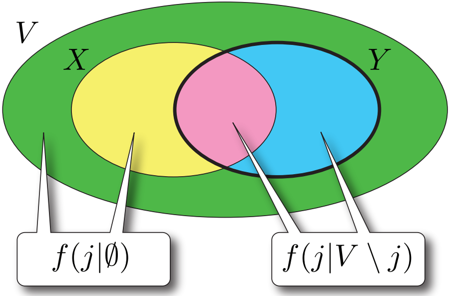


Submodular Supergradients (Iyer et al, 2013)

- These three supergradients (which we call grow, shrink, and bar) are easy to obtain.

Bar:

$$\bar{g}_Y(j) = \begin{cases} f(j|\emptyset) & \text{for } j \notin Y \\ f(j|V \backslash \{j\}) & \text{for } j \in Y \end{cases}$$

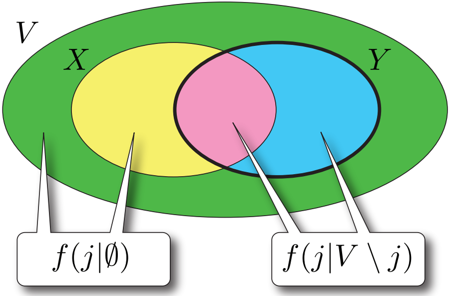


Submodular Supergradients (Iyer et al, 2013)

- These three supergradients (which we call grow, shrink, and bar) are easy to obtain.

Bar:

$$\bar{g}_Y(j) = \begin{cases} f(j|\emptyset) & \text{for } j \notin Y \\ f(j|V\setminus\{j\}) & \text{for } j \in Y \end{cases}$$



- Modular upper bound: $m^{g_Y}(X) = f(Y) + g_Y(X) - g_Y(Y) \leq f(X)$.

Arbitrary functions as difference between submodular funcs.

Theorem

Given an arbitrary set function f, it can be expressed as a difference $f = g - h$ between two polymatroid functions, where both g and h are polymatroidal.

- The semi-gradients above offer a majorization/maximization framework to minimize any function that is naturally expressed as such a difference.

Arbitrary functions as difference between submodular funcs.

Theorem

Given an arbitrary set function f, it can be expressed as a difference $f = g - h$ between two polymatroid functions, where both g and h are polymatroidal.

- The semi-gradients above offer a majorization/maximization framework to minimize any function that is naturally expressed as such a difference.
- E.g., to minimize $f = g - h$, starting with a candidate solution X, repeatedly choose a modular supergradient for g and modular subgradient for h, and perform modular minimization (easy). (see Iyer & Bilmes, 2012).

Applications

- Sensor placement with submodular costs. I.e., let V be a set of possible sensor locations, $f(A) = I(X_A; X_{V \setminus A})$ measures the quality of a subset A of placed sensors, and $c(A)$ the submodular cost. We have $\min_A f(A) - \lambda c(A)$.
- Discriminatively structured graphical models, EAR measure $I(X_A; X_{V \setminus A}) - I(X_A; X_{V \setminus A} | C)$, and synergy in neuroscience.
- Feature selection: a problem of maximizing $I(X_A; C) - \lambda c(A) = H(X_A) - [H(X_A | C) + \lambda c(A)]$, the difference between two submodular functions, where H is the entropy and c is a feature cost function.
- Graphical Model Inference. Finding x that maximizes $p(x) \propto \exp(-v(x))$ where $x \in \{0, 1\}^n$ and v is a pseudo-Boolean function. When v is non-submodular, it can be represented as a difference between submodular functions.

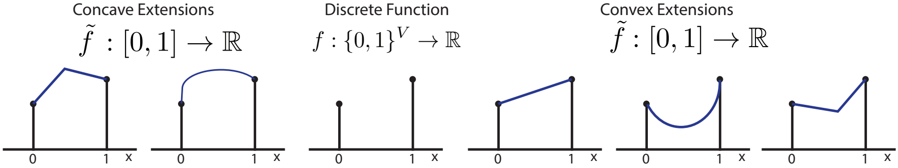
Outline

Continuous Extensions of Discrete Set Functions

- Any function $f : 2^V \to \mathbb{R}$ (equivalently $f : \{0,1\}^V \to \mathbb{R}$) can be extended to a continuous function $\tilde{f} : [0,1]^V \to \mathbb{R}$.

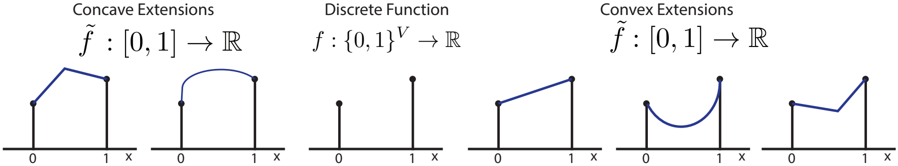
Continuous Extensions of Discrete Set Functions

- Any function $f : 2^V \to \mathbb{R}$ (equivalently $f : \{0,1\}^V \to \mathbb{R}$) can be extended to a continuous function $\tilde{f} : [0,1]^V \to \mathbb{R}$.
- In fact, any such discrete function defined on the vertices of the n-D hypercube $\{0,1\}^n$ has a variety of both convex and concave extensions tight at the vertices (Crama & Hammer). Example $n = 1$,



Continuous Extensions of Discrete Set Functions

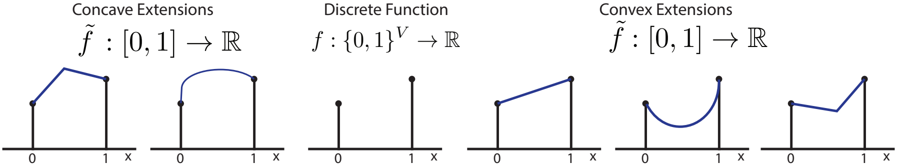
- Any function $f : 2^V \to \mathbb{R}$ (equivalently $f : \{0,1\}^V \to \mathbb{R}$) can be extended to a continuous function $\tilde{f} : [0,1]^V \to \mathbb{R}$.
- In fact, any such discrete function defined on the vertices of the n-D hypercube $\{0,1\}^n$ has a variety of both convex and concave extensions tight at the vertices (Crama & Hammer). Example $n = 1$,



- Since there are an exponential number of vertices $\{0,1\}^n$, important questions regarding such extensions is:

Continuous Extensions of Discrete Set Functions

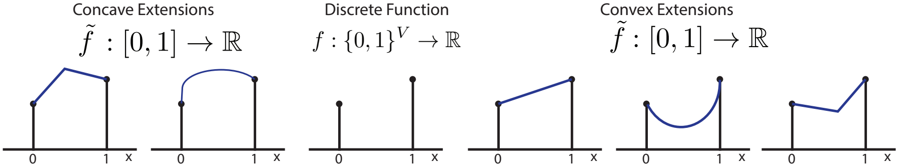
- Any function $f : 2^V \to \mathbb{R}$ (equivalently $f : \{0,1\}^V \to \mathbb{R}$) can be extended to a continuous function $\tilde{f} : [0,1]^V \to \mathbb{R}$.
- In fact, any such discrete function defined on the vertices of the n-D hypercube $\{0,1\}^n$ has a variety of both convex and concave extensions tight at the vertices (Crama & Hammer). Example $n = 1$,



- Since there are an exponential number of vertices $\{0,1\}^n$, important questions regarding such extensions is:
  1. When are they computationally feasible to obtain or estimate?

Continuous Extensions of Discrete Set Functions

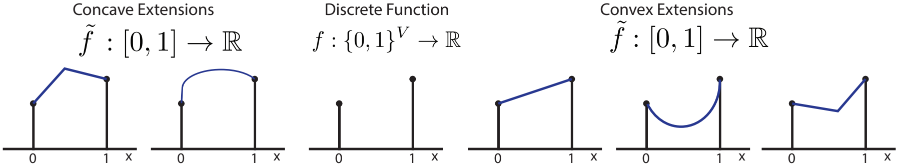
- Any function $f : 2^V \to \mathbb{R}$ (equivalently $f : \{0,1\}^V \to \mathbb{R}$) can be extended to a continuous function $\tilde{f} : [0,1]^V \to \mathbb{R}$.
- In fact, any such discrete function defined on the vertices of the n-D hypercube $\{0,1\}^n$ has a variety of both convex and concave extensions tight at the vertices (Crama & Hammer). Example $n = 1$,



- Since there are an exponential number of vertices $\{0,1\}^n$, important questions regarding such extensions is:
  1. When are they computationally feasible to obtain or estimate?
  2. When do they have nice mathematical properties?

Continuous Extensions of Discrete Set Functions

- Any function $f : 2^V \to \mathbb{R}$ (equivalently $f : \{0,1\}^V \to \mathbb{R}$) can be extended to a continuous function $\tilde{f} : [0,1]^V \to \mathbb{R}$.
- In fact, any such discrete function defined on the vertices of the n-D hypercube $\{0,1\}^n$ has a variety of both convex and concave extensions tight at the vertices (Crama & Hammer). Example $n = 1$,



- Since there are an exponential number of vertices $\{0,1\}^n$, important questions regarding such extensions is:
  1. When are they computationally feasible to obtain or estimate?
  2. When do they have nice mathematical properties?
  3. When are they useful for something practical?

A continuous extension of f

- Given a submodular function f, a $w \in \mathbb{R}^V$, define chain $V_i = \{v_1, v_2, \dots, v_i\}$ based on w sorted in decreasing order. Then Edmonds's greedy algorithm gives us:

$$\tilde{f}(w)$$

A continuous extension of f

- Given a submodular function f, a $w \in \mathbb{R}^V$, define chain $V_i = \{v_1, v_2, \dots, v_i\}$ based on w sorted in decreasing order. Then Edmonds's greedy algorithm gives us:

$$\tilde{f}(w) = \max(wx : x \in P_f) \tag{74}$$

A continuous extension of f

- Given a submodular function f, a $w \in \mathbb{R}^V$, define chain $V_i = \{v_1, v_2, \dots, v_i\}$ based on w sorted in decreasing order. Then Edmonds's greedy algorithm gives us:

$$\tilde{f}(w) = \max(wx : x \in P_f) \tag{74}$$

$$= \sum_{i=1}^{m} w(v_i) f(v_i | V_{i-1}) \tag{75}$$

A continuous extension of f

- Given a submodular function f, a $w \in \mathbb{R}^V$, define chain $V_i = \{v_1, v_2, \ldots, v_i\}$ based on w sorted in decreasing order. Then Edmonds's greedy algorithm gives us:

$$\tilde{f}(w) = \max(wx : x \in P_f) \tag{74}$$

$$= \sum_{i=1}^{m} w(v_i) f(v_i | V_{i-1}) \tag{75}$$

$$= \sum_{i=1}^{m} w(v_i)(f(V_i) - f(V_{i-1})) \tag{76}$$

A continuous extension of f

- Given a submodular function f, a $w \in \mathbb{R}^V$, define chain $V_i = \{v_1, v_2, \dots, v_i\}$ based on w sorted in decreasing order. Then Edmonds's greedy algorithm gives us:

$$\tilde{f}(w) = \max(wx : x \in P_f) \tag{74}$$

$$= \sum_{i=1}^{m} w(v_i) f(v_i | V_{i-1}) \tag{75}$$

$$= \sum_{i=1}^{m} w(v_i)(f(V_i) - f(V_{i-1})) \tag{76}$$

$$= w(v_m) f(V_m) + \sum_{i=1}^{m-1} (w(v_i) - w(v_{i+1})) f(V_i) \tag{77}$$

A continuous extension of f

- Definition of the continuous extension, once again:

$$\tilde{f}(w) = \max(wx : x \in P_f) \tag{78}$$

A continuous extension of f

- Definition of the continuous extension, once again:

$$\tilde{f}(w) = \max(wx : x \in P_f) \tag{78}$$

- Therefore, if f is a submodular function, we can write

$$\tilde{f}(w)$$

A continuous extension of f

- Definition of the continuous extension, once again:

$$\tilde{f}(w) = \max(wx : x \in P_f) \tag{78}$$

- Therefore, if f is a submodular function, we can write

$$\tilde{f}(w) = w(v_m)f(V_m) + \sum_{i=1}^{m-1}(w(v_i) - w(v_{i+1}))f(V_i) \tag{79}$$

A continuous extension of f

- Definition of the continuous extension, once again:

$$\tilde{f}(w) = \max(wx : x \in P_f) \tag{78}$$

- Therefore, if f is a submodular function, we can write

$$\tilde{f}(w) = w(v_m)f(V_m) + \sum_{i=1}^{m-1}(w(v_i) - w(v_{i+1}))f(V_i) \tag{79}$$

$$= \sum_{i=1}^{m} \lambda_i f(V_i) \tag{80}$$

A continuous extension of f

- Definition of the continuous extension, once again:

$$\tilde{f}(w) = \max(wx : x \in P_f) \tag{78}$$

- Therefore, if f is a submodular function, we can write

$$\tilde{f}(w) = w(v_m)f(V_m) + \sum_{i=1}^{m-1}(w(v_i) - w(v_{i+1}))f(V_i) \tag{79}$$

$$= \sum_{i=1}^{m} \lambda_i f(V_i) \tag{80}$$

where $\lambda_m = w(v_m)$ and otherwise $\lambda_i = w(v_i) - w(v_{i+1})$, where the elements are sorted according to w as before.

A continuous extension of f

- Definition of the continuous extension, once again:

$$\tilde{f}(w) = \max(wx : x \in P_f) \tag{78}$$

- Therefore, if f is a submodular function, we can write

$$\tilde{f}(w) = w(v_m)f(V_m) + \sum_{i=1}^{m-1}(w(v_i) - w(v_{i+1}))f(V_i) \tag{79}$$

$$= \sum_{i=1}^{m} \lambda_i f(V_i) \tag{80}$$

where $\lambda_m = w(v_m)$ and otherwise $\lambda_i = w(v_i) - w(v_{i+1})$, where the elements are sorted according to w as before.

- From convex analysis, we know $\tilde{f}(w) = \max(wx : x \in P)$ is always convex in w for any set $P \subseteq R^V$, since it is the maximum of a set of linear functions (true even when f is not submodular or P is not a convex set).

An extension of f

- But, for any $f : 2^V \to \mathbb{R}$, even non-submodular f, we can define an extension in this way, with

$$\tilde{f}(w) = \sum_{i=1}^{m} \lambda_i f(V_i) \tag{81}$$

with the $V_i = \{v_1, \ldots, v_i\}$'s defined based on sorted descending order of w as in $w(v_1) \geq w(v_2) \geq \cdots \geq w(v_m)$, and where

$$\text{for } i \in \{1, \ldots, m\}, \quad \lambda_i = \begin{cases} w(v_i) - w(v_{i+1}) & \text{if } i < m \\ w(v_m) & \text{if } i = m \end{cases} \tag{82}$$

so that $w = \sum_{i=1}^{m} \lambda_i \mathbf{1}_{V_i}$

An extension of f

- But, for any $f : 2^V \to \mathbb{R}$, even non-submodular f, we can define an extension in this way, with

$$\tilde{f}(w) = \sum_{i=1}^{m} \lambda_i f(V_i) \tag{81}$$

with the $V_i = \{v_1, \dots, v_i\}$'s defined based on sorted descending order of w as in $w(v_1) \geq w(v_2) \geq \dots \geq w(v_m)$, and where

$$\text{for } i \in \{1, \dots, m\}, \quad \lambda_i = \begin{cases} w(v_i) - w(v_{i+1}) & \text{if } i < m \\ w(v_m) & \text{if } i = m \end{cases} \tag{82}$$

so that $w = \sum_{i=1}^m \lambda_i \mathbf{1}_{V_i}$

- Note that $w = \sum_{i=1}^m \lambda_i \mathbf{1}_{V_i}$ is an interpolation of certain vertices of the hypercube, and that $\tilde{f}(w) = \sum_{i=1}^m \lambda_i f(V_i)$ is the corresponding interpolation of the values of f at sets corresponding to each hypercube vertex.

Lovász Extension, Submodularity and Convexity

Lovász proved the following important theorem.

Theorem

A function $f : 2^V \to \mathbb{R}$ is submodular iff its its continuous extension defined above as $\tilde{f}(w) = \sum_{i=1}^m \lambda_i f(V_i)$ with $w = \sum_{i=1}^m \lambda_i \mathbf{1}_{V_i}$ is a convex function in $\mathbb{R}^V$.

Lovász extension vs. vector rank function

- Recall vector rank, relative to P_f, defined as:

$$\text{rank}(w) = \max\left(y(V) : y \leq w, y \in P_f\right) \tag{83}$$

where $y \leq w$ is means componentwise inequality ($y_i \leq x_i, \forall i$).

Lovász extension vs. vector rank function

- Recall vector rank, relative to P_f, defined as:

$$\operatorname{rank}(w) = \max\left(y(V) : y \leq w, y \in P_f\right) \tag{83}$$

where $y \leq w$ is means componentwise inequality $(y_i \leq x_i, \forall i)$.

- Lovász extension, $\tilde{f} : [0,1]^V \to \mathbb{R}$:

$$\tilde{f}(w) = \max(w^\intercal x : x \in P_f) \tag{84}$$

Lovász extension vs. vector rank function

- Recall vector rank, relative to P_f, defined as:

$$\text{rank}(w) = \max\left(y(V) : y \leq w, y \in P_f\right) \tag{83}$$

where $y \leq w$ is means componentwise inequality ($y_i \leq x_i, \forall i$).

- Lovász extension, $\tilde{f} : [0,1]^V \to \mathbb{R}$:

$$\tilde{f}(w) = \max(w^\intercal x : x \in P_f) \tag{84}$$

- Both are "submodular" in a sense that $\tilde{f}(a) + \tilde{f}(b) \geq \tilde{f}(a \vee b) + \tilde{f}(a \wedge b)$.

Lovász extension vs. vector rank function

- Recall vector rank, relative to P_f, defined as:

$$\text{rank}(w) = \max\left(y(V) : y \leq w, y \in P_f\right) \tag{83}$$

  where $y \leq w$ is means componentwise inequality ($y_i \leq x_i, \forall i$).

- Lovász extension, $\tilde{f} : [0,1]^V \to \mathbb{R}$:

$$\tilde{f}(w) = \max(w^\intercal x : x \in P_f) \tag{84}$$

- Both are "submodular" in a sense that $\tilde{f}(a) + \tilde{f}(b) \geq \tilde{f}(a \vee b) + \tilde{f}(a \wedge b)$.
- When P_f is a matroid polytope, $\text{rank}(\mathbf{1}_A) = \tilde{f}(\mathbf{1}_A) = \text{rank}(A)$.

Minimizing $\tilde{f}$ vs. minimizing f

Theorem

Let f be submodular and $\tilde{f}$ be its Lovász extension. Then $\min\{f(A)|A \subseteq V\} = \min_{w\in\{0,1\}^V} \tilde{f}(w) = \min_{w\in[0,1]^V} \tilde{f}(w)$.

- Let $w^* \in \operatorname{argmin}\left\{\tilde{f}(w)|w \in [0,1]^V\right\}$ and let $A^* \in \operatorname{argmin}\{f(A)|A \subseteq V\}$.

Minimizing $\tilde{f}$ vs. minimizing f

Theorem

Let f be submodular and $\tilde{f}$ be its Lovász extension. Then $\min\{f(A)|A \subseteq V\} = \min_{w\in\{0,1\}^V} \tilde{f}(w) = \min_{w\in[0,1]^V} \tilde{f}(w)$.

- Let $w^* \in \operatorname{argmin}\left\{\tilde{f}(w)|w \in [0,1]^V\right\}$ and let $A^* \in \operatorname{argmin}\{f(A)|A \subseteq V\}$.
- Define chain $\{V_i^*\}$ based on descending sort of w^*. Then by greedy evaluation of L.E. we have

$$\tilde{f}(w^*) = \sum_i \lambda_i^* f(V_i^*) = f(A^*) = \min\{f(A)|A \subseteq V\} \tag{85}$$

Minimizing $\tilde{f}$ vs. minimizing f

Theorem

Let f be submodular and $\tilde{f}$ be its Lovász extension. Then $\min\{f(A) | A \subseteq V\} = \min_{w \in \{0,1\}^V} \tilde{f}(w) = \min_{w \in [0,1]^V} \tilde{f}(w)$.

- Let $w^* \in \operatorname{argmin}\left\{\tilde{f}(w) | w \in [0,1]^V\right\}$ and let $A^* \in \operatorname{argmin}\{f(A) | A \subseteq V\}$.
- Define chain $\{V_i^*\}$ based on descending sort of w^*. Then by greedy evaluation of L.E. we have

$$\tilde{f}(w^*) = \sum_i \lambda_i^* f(V_i^*) = f(A^*) = \min\{f(A) | A \subseteq V\} \tag{85}$$

- Then we can show that, for each i s.t. $\lambda_i > 0$,

$$f(V_i^*) = f(A^*) \tag{86}$$

So such $\{V_i^*\}$ are also minimizers.

Max-Min Theorems

Theorem

Let f be a submodular function defined on subsets of V. For any $x \in \mathbb{R}^V$, we have:

$$rank(x) = \max\left(y(V) : y \leq x, y \in P_f\right) = \min\left(x(A) + f(V \setminus A) : A \subseteq V\right) \tag{87}$$

If we take x to be zero, we get:

Corollary

Let f be a submodular function defined on subsets of V. $x \in \mathbb{R}^V$, we have:

$$rank(0) = \max\left(y(V) : y \leq 0, y \in P_f\right) = \min\left(f(A) : A \subseteq V\right) \tag{88}$$

Duality of convex minimization of Lovász extension and min-norm point algorithm

- Let f be a submodular function with $\tilde{f}$ it's Lovász extension. Then the following two problems are duals:

$$\underset{w \in \mathbb{R}^V}{\text{minimize}} \; \tilde{f}(w) + \frac{1}{2}\|w\|_2^2 \qquad\qquad \begin{array}{ll} \text{maximize} & -\|x\|_2^2 \\ \text{subject to} & x \in B_f \end{array}$$

  where $B_f = P_f \cap \left\{x \in \mathbb{R}^V : x(V) = f(V)\right\}$ is the base polytope of submodular function f, and $\|x\|_2^2 = \sum_{e \in V} x(e)^2$ is the squared 2-norm.
- Minimum-norm point algorithm (Fujishige-1991, Fujishige-2005, Fujishige-2011, Bach-2013) is essentially an active-set procedure for quadratic programming, and uses Edmonds's greedy algorithm to make it efficient.
- Unknown worst-case running time, although in practice it usually performs quite well.

Other applications of Lovász Extension

- "fast" submodular function minimization, as mentioned above.

Other applications of Lovász Extension

- “fast” submodular function minimization, as mentioned above.
- Structured sparse-encouraging convex norms (Bach-2011, Bach-2012, and Bach-2013), semi-supervised learning, image denoising, etc.

Other applications of Lovász Extension

- "fast" submodular function minimization, as mentioned above.
- Structured sparse-encouraging convex norms (Bach-2011, Bach-2012, and Bach-2013), semi-supervised learning, image denoising, etc.
- E.g., last year's NIPS: Learning scale-free networks (Defazio and Caetano),

Other applications of Lovász Extension

- "fast" submodular function minimization, as mentioned above.
- Structured sparse-encouraging convex norms (Bach-2011, Bach-2012, and Bach-2013), semi-supervised learning, image denoising, etc.
- E.g., last year's NIPS: Learning scale-free networks (Defazio and Caetano),
- Non-linear measures (Denneberg), non-linear aggregation functions (Grabisch et. al), and fuzzy set theory.

Other applications of Lovász Extension

- "fast" submodular function minimization, as mentioned above.
- Structured sparse-encouraging convex norms (Bach-2011, Bach-2012, and Bach-2013), semi-supervised learning, image denoising, etc.
- E.g., last year's NIPS: Learning scale-free networks (Defazio and Caetano),
- Non-linear measures (Denneberg), non-linear aggregation functions (Grabisch et. al), and fuzzy set theory.
- Note, many of the critical properties of the Lovász extension were given by Jack Edmonds in the 1960s. Choquet proposed an identical integral in 1954, and G. Vitali proposed a similar integral in 1925!
  G.Vitali, Sulla definizione di integrale delle funzioni di una variabile, Annali di Matematica Serie IV, Tomo I,(1925), 111-121

Submodular Concave Extension

- Finding a concave extension of a submodular function is NP-hard (Vondrak).

Submodular Concave Extension

- Finding a concave extension of a submodular function is NP-hard (Vondrak).
- However, a useful surrogate is the multi-linear extension.

Submodular Concave Extension

- Finding a concave extension of a submodular function is NP-hard (Vondrak).
- However, a useful surrogate is the multi-linear extension.

Definition

For a set function $f : 2^V \to \mathbb{R}$, define its multilinear extension $F : [0,1]^V \to \mathbb{R}$ by

$$F(x) = \sum_{S \subseteq V} f(S) \prod_{i \in S} x_i \prod_{j \in V \setminus S} (1 - x_j) \tag{90}$$

Submodular Concave Extension

- Finding a concave extension of a submodular function is NP-hard (Vondrak).
- However, a useful surrogate is the multi-linear extension.

Definition

For a set function $f : 2^V \to \mathbb{R}$, define its multilinear extension $F : [0,1]^V \to \mathbb{R}$ by

$$F(x) = \sum_{S \subseteq V} f(S) \prod_{i \in S} x_i \prod_{j \in V \setminus S} (1 - x_j) \tag{90}$$

- Not concave, but still provides useful approximations for many constrained maximization algorithms (e.g., multiple matroid and/or knapsack constraints) via the continuous greedy algorithm followed by rounding.

Submodular Concave Extension

- Finding a concave extension of a submodular function is NP-hard (Vondrak).
- However, a useful surrogate is the multi-linear extension.

Definition

For a set function $f : 2^V \to \mathbb{R}$, define its multilinear extension $F : [0,1]^V \to \mathbb{R}$ by

$$F(x) = \sum_{S \subseteq V} f(S) \prod_{i \in S} x_i \prod_{j \in V \setminus S} (1 - x_j) \tag{90}$$

- Not concave, but still provides useful approximations for many constrained maximization algorithms (e.g., multiple matroid and/or knapsack constraints) via the continuous greedy algorithm followed by rounding.
- Often has to be approximated.

Outline

Submodular: Concave? Convex? Neither? Both?

- Are submodular functions more like convex or more like concave functions?

Submodular is like Concave

- **Convex 1:** Like convex functions, submodular functions can be minimized efficiently (polynomial time).

Submodular is like Concave

- **Convex 1:** Like convex functions, submodular functions can be minimized efficiently (polynomial time).
- **Convex 2:** The Lovász extension of a discrete set function is convex iff the set function is submodular.

Submodular is like Concave

- **Convex 3:** Frank's discrete separation theorem: Let $f : 2^V \to \mathbb{R}$ be a submodular function and $g : 2^V \to \mathbb{R}$ be a supermodular function such that for all $A \subseteq V$,

$$g(A) \leq f(A) \tag{91}$$

Then there exists modular function $x \in \mathbb{R}^V$ such that for all $A \subseteq V$:

$$g(A) \leq x(A) \leq f(A) \tag{92}$$

Submodular is like Concave

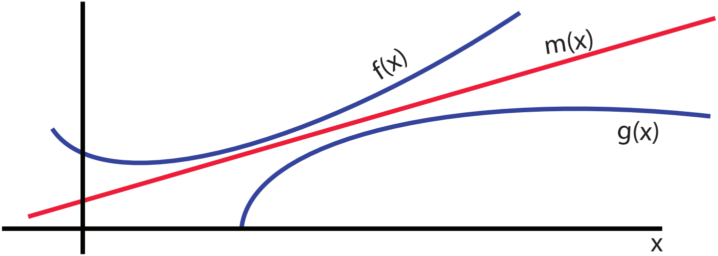
- **Convex 3:** Frank's discrete separation theorem: Let $f : 2^V \to \mathbb{R}$ be a submodular function and $g : 2^V \to \mathbb{R}$ be a supermodular function such that for all $A \subseteq V$,

$$g(A) \leq f(A) \tag{91}$$

Then there exists modular function $x \in \mathbb{R}^V$ such that for all $A \subseteq V$:

$$g(A) \leq x(A) \leq f(A) \tag{92}$$

- Compare to convex/concave case.



Submodular is like Concave

- **Convex 4:** Set of minimizers of a convex function is a convex set. Set of minimizers of a submodular function is a lattice. I.e., if $A, B \in \operatorname{argmin}_{A \subseteq V} f(A)$ then $A \cup B \in \operatorname{argmin}_{A \subseteq V} f(A)$ and $A \cap B \in \operatorname{argmin}_{A \subseteq V} f(A)$

Submodular is like Concave

- **Convex 4:** Set of minimizers of a convex function is a convex set. Set of minimizers of a submodular function is a lattice. I.e., if $A, B \in \operatorname{argmin}_{A \subseteq V} f(A)$ then $A \cup B \in \operatorname{argmin}_{A \subseteq V} f(A)$ and $A \cap B \in \operatorname{argmin}_{A \subseteq V} f(A)$
- **Convex 5:** Submodular functions have subdifferentials and subgradients tight at any point.

Submodularity and Concave

- **Concave 1:** A function is submodular if for all $X \subseteq V$ and $j, k \in V$

$$f(X + j) + f(X + k) \geq f(X + j + k) + f(X) \tag{93}$$

Submodularity and Concave

- **Concave 1:** A function is submodular if for all $X \subseteq V$ and $j, k \in V$

$$f(X + j) + f(X + k) \geq f(X + j + k) + f(X) \tag{93}$$

- With the gain defined as $\nabla_j(X) = f(X + j) - f(X)$, seen as a form of discrete gradient, this 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:

$$\nabla_j \nabla_k f(X) \leq 0 \tag{94}$$

Submodularity and Concave

- **Concave 1:** A function is submodular if for all $X \subseteq V$ and $j, k \in V$
$$f(X + j) + f(X + k) \geq f(X + j + k) + f(X) \tag{93}$$
- With the gain defined as $\nabla_j(X) = f(X + j) - f(X)$, seen as a form of discrete gradient, this 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:
$$\nabla_j \nabla_k f(X) \leq 0 \tag{94}$$
- **Concave 2:** Recall, Theorem 25: composition $h = f \circ g : 2^V \to \mathbb{R}$ (i.e., $h(S) = g(f(S))$) is nondecreasing submodular, if g is non-decreasing concave and f is nondecreasing submodular.

Submodularity and Concave

- **Concave 1:** A function is submodular if for all $X \subseteq V$ and $j, k \in V$

$$f(X+j) + f(X+k) \geq f(X+j+k) + f(X) \tag{93}$$

- With the gain defined as $\nabla_j(X) = f(X+j) - f(X)$, seen as a form of discrete gradient, this 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:

$$\nabla_j \nabla_k f(X) \leq 0 \tag{94}$$

- **Concave 2:** Recall, Theorem 25: composition $h = f \circ g : 2^V \to \mathbb{R}$ (i.e., $h(S) = g(f(S))$) is nondecreasing submodular, if g is non-decreasing concave and f is nondecreasing submodular.
- **Concave 3:** Submodular functions have superdifferentials and supergradients tight at any point.

Submodularity and Concave

- **Concave 1:** A function is submodular if for all $X \subseteq V$ and $j, k \in V$
$$f(X + j) + f(X + k) \geq f(X + j + k) + f(X) \tag{93}$$
- With the gain defined as $\nabla_j(X) = f(X + j) - f(X)$, seen as a form of discrete gradient, this 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:
$$\nabla_j \nabla_k f(X) \leq 0 \tag{94}$$
- **Concave 2:** Recall, Theorem 25: composition $h = f \circ g : 2^V \to \mathbb{R}$ (i.e., $h(S) = g(f(S))$) is nondecreasing submodular, if g is non-decreasing concave and f is nondecreasing submodular.
- **Concave 3:** Submodular functions have superdifferentials and supergradients tight at any point.
- **Concave 4:** Concave maximization solved via local gradient ascent. Submodular maximization is (approximately) solvable via greedy (coordinate-ascent-like) algorithms.

Submodularity and neither Concave nor Convex

- **Neither 1:** Submodular functions have simultaneous sub- and super-gradients, tight at any point.

Submodularity and neither Concave nor Convex

- **Neither 1:** Submodular functions have simultaneous sub- and super-gradients, tight at any point.
- **Neither 2:** Concave functions are closed under min, while submodular functions are not.

Submodularity and neither Concave nor Convex

- **Neither 1:** Submodular functions have simultaneous sub- and super-gradients, tight at any point.
- **Neither 2:** Concave functions are closed under min, while submodular functions are not.
- **Neither 3:** Convex functions are closed under max, while submodular functions are not.

Submodularity and neither Concave nor Convex

- **Neither 1:** Submodular functions have simultaneous sub- and super-gradients, tight at any point.
- **Neither 2:** Concave functions are closed under min, while submodular functions are not.
- **Neither 3:** Convex functions are closed under max, while submodular functions are not.
- **Neither 4:** Convex functions can't, in general, be efficiently or approximately maximized, while submodular functions can be.

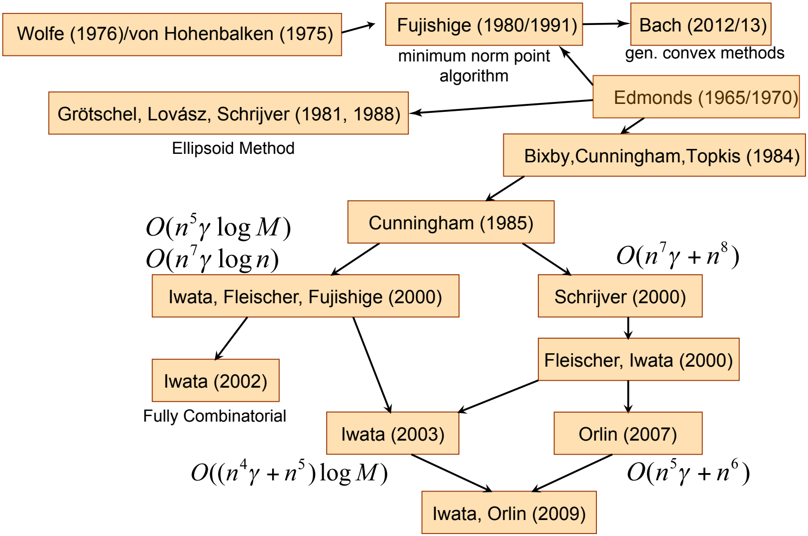
Submodularity and neither Concave nor Convex

- **Neither 1:** Submodular functions have simultaneous sub- and super-gradients, tight at any point.
- **Neither 2:** Concave functions are closed under min, while submodular functions are not.
- **Neither 3:** Convex functions are closed under max, while submodular functions are not.
- **Neither 4:** Convex functions can't, in general, be efficiently or approximately maximized, while submodular functions can be.
- **Neither 5:** Convex functions have local optimality conditions of the form $\nabla_x f(x) = 0$. Analogous submodular function semi-gradient condition $m(X) = 0$ offers no such guarantee (for neither maximization nor minimization).

Outline

SFM Summary (modified from S. Iwata's slides)

General Submodular Function Minimization



Constrained Submodular Minimization

- Constrained submodular minimization

$$\min_{A \in \mathcal{C}} f(A) \tag{95}$$

Constrained Submodular Minimization

- Constrained submodular minimization

$$\min_{A \in \mathcal{C}} f(A) \tag{95}$$

- $\mathcal{C}$ can be paths, matchings, or spanning trees (Goel et. al.), cover constraints (Iwata & Nagano), cuts (Jegelka & Bilmes), or cardinality lower bounds (Svitkina & Fleischer).

Constrained Submodular Minimization

- Constrained submodular minimization

$$\min_{A \in \mathcal{C}} f(A) \tag{95}$$

- $\mathcal{C}$ can be paths, matchings, or spanning trees (Goel et. al.), cover constraints (Iwata & Nagano), cuts (Jegelka & Bilmes), or cardinality lower bounds (Svitkina & Fleischer).
- Some cardinality constraints can be obtained via the min-norm algorithm (Nagano & Kawahara, 2013).

Constrained Submodular Minimization

- Constrained submodular minimization

$$\min_{A \in \mathcal{C}} f(A) \tag{95}$$

- $\mathcal{C}$ can be paths, matchings, or spanning trees (Goel et. al.), cover constraints (Iwata & Nagano), cuts (Jegelka & Bilmes), or cardinality lower bounds (Svitkina & Fleischer).
- Some cardinality constraints can be obtained via the min-norm algorithm (Nagano & Kawahara, 2013).
- Other forms of constraints are “easy” (e.g., certain lattices, odd/even sets (see McCormick's SFM tutorial paper).

Constrained Submodular Minimization

- Constrained submodular minimization

$$\min_{A \in \mathcal{C}} f(A) \tag{95}$$

- $\mathcal{C}$ can be paths, matchings, or spanning trees (Goel et. al.), cover constraints (Iwata & Nagano), cuts (Jegelka & Bilmes), or cardinality lower bounds (Svitkina & Fleischer).
- Some cardinality constraints can be obtained via the min-norm algorithm (Nagano & Kawahara, 2013).
- Other forms of constraints are "easy" (e.g., certain lattices, odd/even sets (see McCormick's SFM tutorial paper).
- In general, many constraints make the problem NP-hard although approximation guarantees are possible (although often hardness is things like $\Omega(n)$ or $\Omega(n^{2/3})$).

Constrained Submodular Minimization

- Constrained submodular minimization

$$\min_{A \in \mathcal{C}} f(A) \tag{95}$$

- $\mathcal{C}$ can be paths, matchings, or spanning trees (Goel et. al.), cover constraints (Iwata & Nagano), cuts (Jegelka & Bilmes), or cardinality lower bounds (Svitkina & Fleischer).
- Some cardinality constraints can be obtained via the min-norm algorithm (Nagano & Kawahara, 2013).
- Other forms of constraints are "easy" (e.g., certain lattices, odd/even sets (see McCormick's SFM tutorial paper).
- In general, many constraints make the problem NP-hard although approximation guarantees are possible (although often hardness is things like $\Omega(n)$ or $\Omega(n^{2/3})$).
- Other forms of constraints: $\mathcal{C} = \{A \subseteq V : g(A) \geq \alpha\}$ for some other submodular function g. This is studied for the first time here at NIPS-2013 (see Saturday talk, Iyer & Bilmes, NIPS 2013).

Submodular Maximization: Unconstrained

- In general, NP-hard.

Submodular Maximization: Unconstrained

- In general, NP-hard.
- The greedy algorithm for monotone submodular maximization:

Algorithm 2: The Greedy Algorithm

Set $S_0 \leftarrow \emptyset$;
for $i \leftarrow 0 \ldots |V| - 1$ **do**
  Choose v_i as follows: $v_i = \left\{ \operatorname{argmax}_{v \in V \setminus S_i} f(S_i \cup \{v\}) \right\}$;
  Set $S_{i+1} \leftarrow S_i \cup \{v_i\}$;

Submodular Maximization: Unconstrained

- In general, NP-hard.
- The greedy algorithm for monotone submodular maximization:

Algorithm 3: The Greedy Algorithm

Set $S_0 \leftarrow \emptyset$;
for $i \leftarrow 0 \ldots |V| - 1$ **do**
  Choose v_i as follows: $v_i = \left\{ \operatorname{argmax}_{v \in V \setminus S_i} f(S_i \cup \{v\}) \right\}$;
  Set $S_{i+1} \leftarrow S_i \cup \{v_i\}$;

- has a strong guarantee:

Submodular Maximization: Unconstrained

- In general, NP-hard.
- The greedy algorithm for monotone submodular maximization:

Algorithm 4: The Greedy Algorithm

Set $S_0 \leftarrow \emptyset$;
for $i \leftarrow 0 \dots |V| - 1$ **do**
  Choose v_i as follows: $v_i = \left\{ \operatorname{argmax}_{v \in V \setminus S_i} f(S_i \cup \{v\}) \right\}$;
  Set $S_{i+1} \leftarrow S_i \cup \{v_i\}$;

- has a strong guarantee:

Theorem

Given a polymatroid function f, the above greedy algorithm returns sets S_i such that for each i we have $f(S_i) \geq (1 - 1/e) \max_{|S| \leq i} f(S)$.

Submodular Max Summary - from J. Vondrak

Monotone Maximization

Constraint	Approximation	Hardness	Technique
$\|S\| \leq k$	$1 - 1/e$	$1 - 1/e$	greedy
matroid	$1 - 1/e$	$1 - 1/e$	multilinear ext.
$O(1)$ knapsacks	$1 - 1/e$	$1 - 1/e$	multilinear ext.
k matroids	$k + \epsilon$	$k/\log k$	local search
k matroids and $O(1)$ knapsacks	$O(k)$	$k/\log k$	multilinear ext.

Nonmonotone Maximization

Constraint	Approximation	Hardness	Technique
Unconstrained	$1/2$	$1/2$	combinatorial
matroid	$1/e$	0.48	multilinear ext.
$O(1)$ knapsacks	$1/e$	0.49	multilinear ext.
k matroids	$k + O(1)$	$k/\log k$	local search
k matroids and $O(1)$ knapsacks	$O(k)$	$k/\log k$	multilinear ext.

Outline

A submodular function as a parameter

- In some cases, it may be useful to view a submodular function $f : 2^V \to \mathbb{R}$ as a input "parameter" to a machine learning algorithm.

A submodular function as a parameter

- In some cases, it may be useful to view a submodular function $f : 2^V \to \mathbb{R}$ as a input "parameter" to a machine learning algorithm.
- Hence, it is imperative in the ML community to develop ways to learn or approximately learn such submodular parameterizations.

A submodular function as a parameter

- In some cases, it may be useful to view a submodular function $f : 2^V \to \mathbb{R}$ as a input "parameter" to a machine learning algorithm.
- Hence, it is imperative in the ML community to develop ways to learn or approximately learn such submodular parameterizations.
- Ex: Structured sparsity-encouraging convex norm (Bach): i.e., a submodular function f, via its Lovász extension $\tilde{f}$, gives us a norm

$$\|w\|_f = \tilde{f}(|w|) \tag{96}$$

A submodular function as a parameter

- In some cases, it may be useful to view a submodular function $f : 2^V \to \mathbb{R}$ as a input "parameter" to a machine learning algorithm.
- Hence, it is imperative in the ML community to develop ways to learn or approximately learn such submodular parameterizations.
- Ex: Structured sparsity-encouraging convex norm (Bach): i.e., a submodular function f, via its Lovász extension $\tilde{f}$, gives us a norm

$$\|w\|_f = \tilde{f}(|w|) \tag{96}$$

- So finding a desirable norm is equivalent to finding a desirable submodular function.

Graphical Models vs. log-supermodular distributions

- Consider family of distributions $p : \{0,1\}^V \to [0,1]$ of the form:

$$p(x) = \frac{1}{Z}\exp(f(x)) \tag{97}$$

Graphical Models vs. log-supermodular distributions

- Consider family of distributions $p : \{0,1\}^V \to [0,1]$ of the form:

$$p(x) = \frac{1}{Z}\exp(f(x)) \tag{97}$$

- Graphical models: $f(x) = \sum_{c \in \mathcal{C}} f_c(x_c)$ where $\mathcal{C}$ are a set of cliques.

Graphical Models vs. log-supermodular distributions

- Consider family of distributions $p : \{0,1\}^V \to [0,1]$ of the form:

$$p(x) = \frac{1}{Z} \exp(f(x)) \tag{97}$$

- Graphical models: $f(x) = \sum_{c \in \mathcal{C}} f_c(x_c)$ where $\mathcal{C}$ are a set of cliques.
- If $-f$ is supermodular, MAP assignment is a submodular minimization problem. Typical example:

$$p(x) = \frac{1}{Z} \exp(-f(x) + m(x)) \tag{98}$$

where f is submodular "energy" (often a graph-cut problem) and m is modular (unaries). Common in computer vision.

Graphical Models vs. log-supermodular distributions

- Consider family of distributions $p : \{0,1\}^V \to [0,1]$ of the form:

$$p(x) = \frac{1}{Z}\exp(f(x)) \tag{97}$$

- Graphical models: $f(x) = \sum_{c \in \mathcal{C}} f_c(x_c)$ where $\mathcal{C}$ are a set of cliques.
- If $-f$ is supermodular, MAP assignment is a submodular minimization problem. Typical example:

$$p(x) = \frac{1}{Z}\exp(-f(x) + m(x)) \tag{98}$$

  where f is submodular "energy" (often a graph-cut problem) and m is modular (unaries). Common in computer vision.
- Complexity is polynomial regardless of the tree-width of f — submodularity is anti-graphical.

Graphical Models vs. log-supermodular distributions

- Consider family of distributions $p : \{0,1\}^V \to [0,1]$ of the form:

$$p(x) = \frac{1}{Z}\exp(f(x)) \tag{97}$$

- Graphical models: $f(x) = \sum_{c \in \mathcal{C}} f_c(x_c)$ where $\mathcal{C}$ are a set of cliques.
- If $-f$ is supermodular, MAP assignment is a submodular minimization problem. Typical example:

$$p(x) = \frac{1}{Z}\exp(-f(x) + m(x)) \tag{98}$$

  where f is submodular "energy" (often a graph-cut problem) and m is modular (unaries). Common in computer vision.
- Complexity is polynomial regardless of the tree-width of f — submodularity is anti-graphical.
- Log-supermodular distributions, since $\log p(x)$ is a supermodular function.

Graphical Models vs. log-submodular distributions

- On the other hand, with

$$p(x) = \frac{1}{Z} \exp(f(x)) \tag{99}$$

where f is submodular then p is log-submodular.

Graphical Models vs. log-submodular distributions

- On the other hand, with

$$p(x) = \frac{1}{Z} \exp(f(x)) \tag{99}$$

where f is submodular then p is log-submodular.

- Example: $f(x) = \log \det(\mathbf{C}_X)$ where X is the set corresponding to binary vector x (i.e., $x = \mathbf{1}_X$) and C is the sub-matrix with rows/columns selected by X.

Graphical Models vs. log-submodular distributions

- On the other hand, with

$$p(x) = \frac{1}{Z} \exp(f(x)) \tag{99}$$

where f is submodular then p is log-submodular.

- Example: $f(x) = \log\det(\mathbf{C}_X)$ where X is the set corresponding to binary vector x (i.e., $x = \mathbf{1}_X$) and C is the sub-matrix with rows/columns selected by X.
- Hence $p(x)$ is a determinantal point process.

log-supermodular vs. log-submodular distributions

- Log-supermodular: MAP or high-probable assignments should be "regular", "homogeneous", "smooth", "simple". E.g., attractive potentials in computer vision, ferromagnetic Potts model statistical physics.

log-supermodular vs. log-submodular distributions

- Log-supermodular: MAP or high-probable assignments should be "regular", "homogeneous", "smooth", "simple". E.g., attractive potentials in computer vision, ferromagnetic Potts model statistical physics.
- Log-submodular: MAP or high-probable assignments should be "diverse", or "complex", or "covering", like in determinantal point processes.

Other examples: submodular parameterization

- Combinatorial independence, generalized entropy, and “information” or “complexity” functions (seen above).

Other examples: submodular parameterization

- Combinatorial independence, generalized entropy, and "information" or "complexity" functions (seen above).
- Simultaneous batch active-learning/semi-supervised learning (Guillory & Bilmes).

Other examples: submodular parameterization

- Combinatorial independence, generalized entropy, and “information” or “complexity” functions (seen above).
- Simultaneous batch active-learning/semi-supervised learning (Guillory & Bilmes).
- Rank-order based divergences (Submodular Bregman Divergence, and the Lovász-Bregman Divergences) (Iyer & Bilmes, 2013).

Other examples: submodular parameterization

- Combinatorial independence, generalized entropy, and "information" or "complexity" functions (seen above).
- Simultaneous batch active-learning/semi-supervised learning (Guillory & Bilmes).
- Rank-order based divergences (Submodular Bregman Divergence, and the Lovász-Bregman Divergences) (Iyer & Bilmes, 2013).
- Feature and dictionary selection (Krause & Guestrin, Das & Kempe)

Other examples: submodular parameterization

- Combinatorial independence, generalized entropy, and "information" or "complexity" functions (seen above).
- Simultaneous batch active-learning/semi-supervised learning (Guillory & Bilmes).
- Rank-order based divergences (Submodular Bregman Divergence, and the Lovász-Bregman Divergences) (Iyer & Bilmes, 2013).
- Feature and dictionary selection (Krause & Guestrin, Das & Kempe)
- Computer vision (Kolmogorov, Boykov, Kohli, Ladicky, Torr, etc.).

Other examples: submodular parameterization

- Combinatorial independence, generalized entropy, and "information" or "complexity" functions (seen above).
- Simultaneous batch active-learning/semi-supervised learning (Guillory & Bilmes).
- Rank-order based divergences (Submodular Bregman Divergence, and the Lovász-Bregman Divergences) (Iyer & Bilmes, 2013).
- Feature and dictionary selection (Krause & Guestrin, Das & Kempe)
- Computer vision (Kolmogorov, Boykov, Kohli, Ladicky, Torr, etc.).
- Data subset (or core set) selection in machine learning (Lin & Bilmes, Wei & Bilmes). Data summarization, summarizing big redundant data.

Other examples: submodular parameterization

- Combinatorial independence, generalized entropy, and “information” or “complexity” functions (seen above).
- Simultaneous batch active-learning/semi-supervised learning (Guillory & Bilmes).
- Rank-order based divergences (Submodular Bregman Divergence, and the Lovász-Bregman Divergences) (Iyer & Bilmes, 2013).
- Feature and dictionary selection (Krause & Guestrin, Das & Kempe)
- Computer vision (Kolmogorov, Boykov, Kohli, Ladicky, Torr, etc.).
- Data subset (or core set) selection in machine learning (Lin & Bilmes, Wei & Bilmes). Data summarization, summarizing big redundant data.
- Influence determination in social networks (Kempe, Kleinberg, & Tardos)

Outline

Classic References

- Jack Edmonds's paper "Submodular Functions, Matroids, and Certain Polyhedra" from 1970.
- Nemhauser, Wolsey, Fisher, "A Analysis of Approximations for Maximizing Submodular Set Functions-I", 1978
- Lovász's paper, "Submodular functions and convexity", from 1983.

Classic Books

- Fujishige, "Submodular Functions and Optimization", 2005
- Narayanan, "Submodular Functions and Electrical Networks", 1997
- Welsh, "Matroid Theory", 1975.
- Oxley, "Matroid Theory", 1992 (and 2011).
- Lawler, "Combinatorial Optimization: Networks and Matroids", 1976.
- Schrijver, "Combinatorial Optimization", 2003
- Gruenbaum, "Convex Polytopes, 2nd Ed", 2003.

Recent online material with an ML slant

- My class, most proofs for above are given. http://j.ee.washington.edu/~bilmes/classes/ee596a_fall_2012/. Next offered, April 2014.
- Andreas Krause's web page http://submodularity.org.
- Stefanie Jegelka and Andreas Krause's ICML 2013 tutorial http://techtalks.tv/talks/submodularity-in-machine-learning-new-directions-part-i/58125/
- Francis Bach's updated 2013 text. http://hal.archives-ouvertes.fr/docs/00/87/06/09/PDF/submodular_fot_revised_hal.pdf
- Tom McCormick's overview paper on submodular minimization http://people.commerce.ubc.ca/faculty/mccormick/sfmchap8a.pdf
- Georgia Tech's 2012 workshop on submodularity: http://www.arc.gatech.edu/events/arc-submodularity-workshop

The End: Thank you!