Logistics

Review

Cumulative Outstanding Reading

- Read chapter 1 from Fujishige's book.
Announcements, Assignments, and Reminders

- our room (Mueller Hall Room 154) is changed!
- Please do use our discussion board (https://canvas.uw.edu/courses/895956/discussion_topics) for all questions, comments, so that all will benefit from them being answered.
- Weekly Office Hours: Wednesdays, 5:00-5:50, or by skype or google hangout (email me).

Class Road Map - IT-I

L1 (3/31): Motivation, Applications, & Basic Definitions
L2: (4/2): Applications, Basic Definitions, Properties
L3: More examples and properties (e.g., closure properties), and examples, spanning trees
L4: proofs of equivalent definitions, independence, start matroids
L5:
L6:
L7:
L8:
L9:
L10:
L11:
L12:
L13:
L14:
L15:
L16:
L17:
L18:
L19:
L20:

Finals Week: June 9th-13th, 2014.
Submodular Definitions

Definition 3.2.2 (submodular concave)

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

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B)$$  \hspace{1cm} (3.2)

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

Definition 3.2.3 (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:

$$f(A \cup \{v\}) - f(A) \geq f(B \cup \{v\}) - f(B)$$  \hspace{1cm} (3.3)

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

Many Properties

- In the last lecture, we started looking at properties of and gaining intuition about submodular functions.
- We began to see that there were many functions that were submodular, and operations on sets of submodular functions that preserved submodularity.
Some examples form last time

- Coverage functions (either via sets, or via regions in $n$-D space).
- Entropy function (as a function of sets of random variables), symmetric mutual information.
- Many functions based on graphs are either submodular or supermodular, and other functions might not be (e.g., graph strength) but involve submodularity in a critical way.
- Matrix rank - rank of a set of vectors from a set of vector indices.
- Geometric interpretation of $f(A) + f(B) \geq f(A \cup B) + f(A \cap B)$.
- Cost of manufacturing – supply side economies of scale
- Network Externalities – Demand side Economies of Scale
- Social Network Influence
- Information and Summarization - document summarization via sentence selection

The Venn and Art of Submodularity

\[ r(A) + r(B) \geq r(A \cup B) + r(A \cap B) \]
\[ = r(A_r) + 2r(C) + r(B_r) \]
\[ = r(A_r) + r(C) + r(B_r) \]
\[ = r(A \cap B) \]
**Polymatroid rank function**

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

$$f(X) = r(\bigcup_{s \in S} X_s) \quad (3.1)$$

we have that $f$ is submodular, and is known to be a polymatroid rank function.

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

**Spanning trees**

- Let $E$ be a set of edges of some graph $G = (V, E)$, and let $r(S)$ for $S \subseteq E$ be the maximum size (in terms of number of edges) spanning forest in the vertex-induced graph, induced by vertices incident to edges $S$.
- Example: Given $G = (V, E)$, $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$, $E = \{1, 2, \ldots, 12\}$. $S = \{1, 2, 3, 4, 5, 8, 9\} \subseteq E$. Two spanning trees have the same edge count (the rank of $S$).
- Then $r(S)$ is submodular, and is another matrix rank function corresponding to the incidence matrix of the graph.
Let $V$ be a set of information containing elements ($V$ might say be either words, sentences, documents, web pages, or blogs, each $v \in V$ is one element, so $v$ might be a word, a sentence, a document, etc.). The total amount of information in $V$ is measured by a function $f(V)$, and any given subset $S \subseteq V$ measures the amount of information in $S$, given by $f(S)$.

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

So a submodular function would likely be a good model.

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

\[
P_f = \{ x \in \mathbb{R}^E : x(S) \leq f(S), \forall S \subseteq E \} \quad (3.2)
\]

\[
P^+_f = P_f \cap \{ x \in \mathbb{R}^E : x \geq 0 \} \quad (3.3)
\]

\[
B_f = P_f \cap \{ x \in \mathbb{R}^E : x(E) = f(E) \} \quad (3.4)
\]

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

\[
\tilde{f}(c) \triangleq \max \{ c^T x : x \in P_f \} \quad (3.5)
\]

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

- It is common in the literature to use either $E$ or $V$ as the ground set.
- We will follow this inconsistency in the literature and will inconsistently use either $E$ or $V$ as our ground set (hopefully not in the same equation, if so, please point this out).

Notation $\mathbb{R}^E$

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

$$\mathbb{R}^E = \{x = (x_j \in \mathbb{R} : j \in E)\} \quad (3.6)$$

$$\mathbb{R}^E_+ = \{x = (x_j : j \in E) : x \geq 0\} \quad (3.7)$$

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

$$x(A) = \sum_{a \in A} x_a \quad (3.8)$$

Note that $x$ is said to be normalized since $x(\emptyset) = 0$. 
characteristic vectors of sets & modular functions

- Given an $A \subseteq E$, define the vector $1_A \in \mathbb{R}^E_+$ to be
  \[
  1_A(j) = \begin{cases} 
  1 & \text{if } j \in A; \\
  0 & \text{if } j \notin A
  \end{cases}
  \]  
  (3.9)

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

- Thus, given modular function $x \in \mathbb{R}^E$, we can write $x(A)$ in a variety of ways, i.e.,
  \[
  x(A) = x \cdot 1_A = \sum_{i \in A} x(i)
  \]  
  (3.10)

Other Notation: singletons and sets

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

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

Summing Submodular Functions

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

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

is submodular. This follows easily since

$$f(A) + f(B) = f_1(A) + f_2(A) + f_1(B) + f_2(B) \quad (3.12)$$
$$\geq f_1(A \cup B) + f_2(A \cup B) + f_1(A \cap B) + f_2(A \cap B) \quad (3.13)$$
$$= f(A \cup B) + f(A \cap B). \quad (3.14)$$

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

Summing Submodular and Modular Functions

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

\[
    f : 2^E \rightarrow \mathbb{R} \text{ with } f(A) = f_1(A) - m(A) \tag{3.15}
\]

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

\[
    f(A) + f(B) = f_1(A) - m(A) + f_1(B) - m(B) \geq f_1(A \cup B) - m(A \cup B) + f_1(A \cap B) - m(A \cap B) \tag{3.17}
    = f(A \cup B) + f(A \cap B). \tag{3.18}
\]

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

Restricting Submodular Functions

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

\[
    f' : 2^E \rightarrow \mathbb{R} \text{ with } f'(A) = f(A \cap S) \tag{3.19}
\]

is submodular.

**Proof.**

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

\[
    f((A+v) \cap S) - f(A \cap S) \geq f((B+v) \cap S) - f(B \cap S) \tag{3.20}
\]

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

\[
    f(A' + v) - f(A') \geq f(B' + v) - f(B') \tag{3.21}
\]

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

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

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

(3.22)

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

Given $V$, let $C = \{C_1, C_2, \ldots, C_k\}$ be a set of subsets of $V$, and for each $C \in C$, let $f_C : 2^V \rightarrow \mathbb{R}$ be a submodular function. Then

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

(3.23)

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

Max - normalized

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

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

(3.24)

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

Proof.

Consider

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

(3.25)

which follows since we have that

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

(3.26)

and

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

(3.27)
Max

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

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

(3.28)

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

Proof.
The proof is identical to the normalized case.

Facility/Plant Location (uncapacitated)

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

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

(3.4)

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

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

\[
f : 2^E \to \mathbb{R}, \quad \text{where } f(A) = \sum_{i \in V} \max_{j \in A} c_{ij}\tag{3.29}
\]

is submodular.

Proof.

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

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

Log Determinant

- Let \( \Sigma \) be an \( n \times n \) positive definite matrix. Let \( V = \{1, 2, \ldots, n\} \equiv [n] \) be an index set, and for \( A \subseteq V \), let \( \Sigma_A \) be the (square) submatrix of \( \Sigma \) obtained by including only entries in the rows/columns given by \( A \).
- We have that:
  \[
f(A) = \log \det(\Sigma_A) \text{ is submodular.} \tag{3.30}
\]
- The submodularity of the log determinant is crucial for determinantal point processes (DPPs) (defined later in the class).

Proof of submodularity of the logdet function.

Suppose \( X \in \mathbb{R}^n \) is multivariate Gaussian random variable, that is

\[
x \in p(x) = \frac{1}{\sqrt{2\pi |\Sigma|}} \exp \left( -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right) \tag{3.31}
\]

...
Log Determinant

...cont.

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

$$h(X) = \log \sqrt{2\pi e \Sigma} = \log \sqrt{(2\pi e)^n |\Sigma|} \quad (3.32)$$

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

$$f(A) = h(X_A) = \log \sqrt{(2\pi e)^{|A| |\Sigma_A|}} \quad (3.33)$$

Entropy is submodular (conditioning reduces entropy), and moreover

$$f(A) = h(X_A) = m(A) + \frac{1}{2} \log |\Sigma_A| \quad (3.34)$$

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

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

Summary so far

- Summing: if $\alpha_i \geq 0$ and $f_i : 2^V \to \mathbb{R}$ is submodular, then so is $\sum_i \alpha_i f_i$.
- Restrictions: $f'(A) = f(A \cap S)$
- max: $f(A) = \max_{j \in A} c_j$ and facility location.
- Log determinant $f(A) = \log \det(\Sigma_A)$
Concave over non-negative modular

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

$$f(A) = g(m(A))$$  \hspace{1cm} (3.35)

then $f$ is submodular.

**Proof.**

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

$$g(m(A) + m(v)) - g(m(A)) \geq g(m(B) + m(v)) - g(m(B))$$  \hspace{1cm} (3.36)

A form of converse is true as well.

Concave composed with non-negative modular

**Theorem 3.5.1**

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

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

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

$$f(A) = \sum_{i=1}^{K} g_i(m_i(A))$$  \hspace{1cm} (3.37)

- Very large class of functions, including graph cut, bipartite neighborhoods, set cover (Stobbe & Krause).
- However, Vondrak showed that a graphic matroid rank function over $K_4$ (we’ll define this after we define matroids) are not members.
Monotonicity

**Definition 3.6.1**

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

**Definition 3.6.2**

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

**Composition of non-decreasing submodular and non-decreasing concave**

**Theorem 3.6.3**

*Given two functions, one defined on sets*

\[
f : 2^V \to \mathbb{R} \tag{3.38}
\]

*and another continuous valued one:*

\[
g : \mathbb{R} \to \mathbb{R} \tag{3.39}
\]

*the composition formed as \( h = g \circ f : 2^V \to \mathbb{R} \) (defined as \( h(S) = g(f(S)) \)) is nondecreasing submodular, if \( g \) is non-decreasing concave and \( f \) is nondecreasing submodular.*
Monotone difference of two functions

Let \( f \) and \( g \) both be submodular functions on subsets of \( V \) and let \((f - g)(\cdot)\) be either monotone increasing or monotone decreasing. Then \( h : 2^V \rightarrow \mathbb{R} \) defined by

\[
h(A) = \min(f(A), g(A)) \tag{3.40}
\]

is submodular.

**Proof.**

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

\[
f(X) + f(Y) \geq f(X \cup Y) + f(X \cap Y) \tag{3.41}
\]
\[
g(X) + g(Y) \geq g(X \cup Y) + g(X \cap Y), \tag{3.42}
\]

the result (Equation 3.40) follows since

\[
f(X) + f(Y) \geq \min(f(X \cup Y), g(X \cup Y)) + \min(f(X \cap Y), g(X \cap Y)) \tag{3.43}
\]

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

\[
h(X) + h(Y) = f(X) + g(Y) \geq f(X \cup Y) + f(X \cap Y) + g(Y) - f(Y) \tag{3.44}
\]

Assume the case where \( f - g \) is monotone increasing. Hence,

\[
f(X \cup Y) + g(Y) - f(Y) \geq g(X \cup Y) \text{ giving}
\]

\[
h(X) + h(Y) \geq g(X \cup Y) + f(X \cap Y) \geq h(X \cup Y) + h(X \cap Y) \tag{3.45}
\]

What is an easy way to prove the case where \( f - g \) is monotone decreasing?
### Saturation via the \( \min(\cdot) \) function

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

\[
h(A) = \min(k, f(A))
\]  

is submodular.

**Proof.**

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

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

### More on Min - the saturate trick

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

\[
h(A) = \frac{1}{2}(\min(k, f) + \min(k, g))
\]  

then \( h \) is submodular, and \( h(A) \geq k \) if and only if both \( f(A) \geq k \) and \( g(A) \geq k \).

- This can be useful in many applications. Moreover, this is an instance of a submodular surrogate (where we take a non-submodular problem and find a submodular one that can tell us something). We hope to revisit this again later in the quarter.
Arbitrary functions as difference between submodular funcs.

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

**Proof.**

Let $f$ be given and arbitrary, and define:

$$
\alpha \triangleq \min_{X,Y} \left( f(X) + f(Y) - f(X \cup Y) - f(X \cap Y) \right)
$$

(3.48)

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

$$
\beta \triangleq \min_{X,Y} \left( h(X) + h(Y) - h(X \cup Y) - h(X \cap Y) \right).
$$

(3.49)

Strict means that $\beta > 0$.

...cont.

Define $f' : 2^V \rightarrow \mathbb{R}$ as

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

(3.50)

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

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

  \[
  f(A \cup \{j\}) - f(A) \triangleq \rho_j(A) \tag{3.51}
  \]
  \[
  \triangleq \rho_A(j) \tag{3.52}
  \]
  \[
  \triangleq \nabla_jf(A) \tag{3.53}
  \]
  \[
  \triangleq f(\{j\}|A) \tag{3.54}
  \]
  \[
  \triangleq f(j|A) \tag{3.55}
  \]

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

Gain Notation

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

\[
 f(A|B) \triangleq f(A \cup B) - f(B) \tag{3.56}
\]

So when $j$ is any singleton

\[
 f(j|B) = f(\{j\}|B) = f(\{j\} \cup B) - f(B) \tag{3.57}
\]

Note that this is inspired from information theory and the notation used for conditional entropy $H(X_A|X_B) = H(X_A, X_B) - H(X_B)$. 

### Arbitrary function as difference between two polymatroids

- Any submodular function \( g \) can be represented as a sum of a polymatroid (normalized monotone non-decreasing submodular) function \( \bar{g} \) and a modular function \( m_g \).
- Given submodular \( g : 2^V \rightarrow \mathbb{R} \), construct \( \bar{g} : 2^V \rightarrow \mathbb{R} \) as \( \bar{g}(A) = g(A) - \sum_{a \in A} g(a|V \setminus \{a\}) \). Let \( m_g(A) \triangleq \sum_{a \in A} g(a|V \setminus \{a\}) \).
- Then, given arbitrary \( f = g - h \) where \( g \) and \( h \) are submodular,

\[
\begin{align*}
    f &= g - h = \bar{g} + m_g - \bar{h} - m_h \\
    &= \bar{g} - \bar{h} + (m_g - m_h) \\
    &= \bar{g} - \bar{h} + m_{g-h} \\
    &= \bar{g} + m^+_{g-h} - (\bar{h} + (-m_{g-h})^+) 
\end{align*}
\]

where \( m^+ \) is the positive part of modular function \( m \). That is, \( m^+(A) = \sum_{a \in A} m(a)1(m(a) > 0) \).
- But both \( g + m^+_{g-h} \) and \( \bar{h} + (-m_{g-h})^+ \) are polymatroid functions.
- Thus, any function can be expressed as a difference between two polymatroid functions.

### 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 \( f(A) - \lambda c(A) \) as the overall objective.
- **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.