EE595A – Submodular functions, their optimization and applications – Spring 2011

Prof. Jeff Bilmes

University of Washington, Seattle
Department of Electrical Engineering
Spring Quarter, 2011
http://ssli.ee.washington.edu/~bilmes/ee595a_spring_2011/

Lecture 1 - March 30th, 2011
Welcome to the class!

Weekly Office Hours: Wednesdays, 12:30-1:30pm, 10 minutes after class on Wednesdays.
This course will serve as an introduction to submodular functions including methods for their optimization, and how they have been (and can be) applied in many application domains.
Introduction to submodular functions, including definitions, real-world and contrived examples of submodular functions, properties, operations that preserve submodularity, submodular variants and special submodular functions, and computational properties.

Background on submodular functions, including a brief overview of the theory of matroids and lattices.

Polyhedral properties of submodular functions

The Lovász extension of submodular functions. The Choquet integral.

Submodular maximization algorithms under simple constraints, submodular cover problems, greedy algorithms, approximation guarantees.
Rough Outline (cont. II)

- Submodular minimization algorithms, a history of submodular minimization, including both numerical and combinatorial algorithms, computational properties of these algorithms, and descriptions of both known results and currently open problems in this area.

- Submodular flow problems, the principle partition of a submodular function and its variants.

- Constrained optimization problems with submodular functions, including maximization and minimization problems with various constraints. An overview of recent problems addressed in the community.

- Applications of submodularity in computer vision, constraint satisfaction, game theory, information theory, norms, natural language processing, graphical models, and machine learning
Facts about the class

- Prerequisites: ideally knowledge in probability, statistics, convex optimization, and combinatorial optimization these will be reviewed as necessary. The course is open to students in all UW departments. Any questions, please contact me.
- Text: We will be drawing from the book by Satoru Fujishige entitled "Submodular Functions and Optimization" 2nd Edition, 2005, but we will also be reading research papers that will be posted here on this web page, especially for some of the application areas.
- Grades and Assignments: Grades will be based on a combination of a final project (35%), homeworks (35%), and the take home midterm exam (30%). There will be between 3-4 homeworks during the quarter.
- Final project: The final project will consist of a 4-page paper (conference style) and a final project presentation. The project must involve using/dealing mathematically with submodularity in some way or another.
Facts about the class

- Homework/midterm must be submitted electronically using our web page (see http://ssli.ee.washington.edu/~bilmes/ee595a_spring_2011/). PDF submissions only please.
- Lecture slides - are being prepared as we speak. I will try to have them up on the web page the night before each class. I will not only draw from the book but other sources which will be listed at the end of each set of slides.
- Our currently scheduled final presentation is Monday, June 06, 2011, 830-1020, MUE 155. Perhaps this is a bit early on a Monday. Finals week is June 6-10th, so we’ll find a time later this week to do it.
Submodular Motivation

- Given a set of objects $V = \{v_1, \ldots, v_n\}$ and a function $f : 2^V \rightarrow \mathbb{R}$ that returns a real value for any subset $S \subseteq V$.
- Suppose we are interested in finding the subset that either maximizes or minimizes the function, e.g., $\text{argmax}_{S \subseteq V} f(S)$, possibly subject to some constraints.
- In general, this problem has exponential time complexity.
- Example: $f$ might correspond to the value (e.g., information gain) of a set of sensor locations in an environment, and we wish to find the best set $S \subseteq V$ of sensors locations given a fixed upper limit on the number of sensors $|S|$.
- In many cases (such as above) $f$ has properties that make its optimization tractable to either exactly or approximately compute.
- One such property is submodularity.
Submodular Definitions

**Definition (submodular)**

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

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B) \quad (1)$$

An alternate and equivalent definition is:

**Definition (diminishing returns)**

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

$$f(A \cup \{v\}) - f(A) \geq f(B \cup \{v\}) - f(B) \quad (2)$$

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 $B$. 
Subadditive Definitions

Definition (subadditive)

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

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

(3)

This means that the “whole” is less than the sum of the parts.
Supermodular Definitions

**Definition (supermodular)**

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

$$f(A) + f(B) \leq f(A \cup B) + f(A \cap B) \quad (4)$$

An alternate and equivalent definition is:

**Definition (increasing returns)**

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

$$f(A \cup \{v\}) - f(A) \leq f(B \cup \{v\}) - f(B) \quad (5)$$

The incremental “value”, “gain”, or “cost” of $v$ increases as the context in which $v$ is considered grows from $A$ to $B$. 
Superadditive Definitions

Definition (superadditive)

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

\[
f(A) + f(B) \leq f(A \cup B)
\]  

(6)

This means that the “whole” is greater than the sum of the parts.
Supermodular/Superadditive Anecdote

From David Brooks, NYT's column, March 28th, 2011 on “Tools for Thinking”. In response to Steven Pinker (Harvard) asking a number of people “What scientific concept would improve everybody’s cognitive toolkit?”

Emergent systems are ones in which many different elements interact. The pattern of interaction then produces a new element that is greater than the sum of the parts, which then exercises a top-down influence on the constituent elements.
Definition (modular)

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

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

$$f(A) + f(B) = f(A \cap B) + f(A \cup B) \quad (7)$$

Modular functions have no interaction, and have value based only on singleton values.

**Proposition**

*If $f$ is modular, it may be written as*

$$f(A) = f(\emptyset) + \sum_{a \in A} \left( f(\{a\}) - f(\emptyset) \right) \quad (8)$$
Proof.

We inductively construct the value for $A = \{a_1, a_2, \ldots, a_k\}$.

$$f(a_1) + f(a_2) = f(a_1, a_2) + f(\emptyset) \quad (9)$$
implies
$$f(a_1, a_2) = f(a_1) - f(\emptyset) + f(a_2) - f(\emptyset) + f(\emptyset) \quad (10)$$

then

$$f(a_1, a_2) + f(a_3) = f(a_1, a_2, a_3) + f(\emptyset) \quad (11)$$
implies
$$f(a_1, a_2, a_3) = f(a_1, a_2) - f(\emptyset) + f(a_3) - f(\emptyset) + f(\emptyset) \quad (12)$$

$$= f(\emptyset) + \sum_{i=1}^{3} f(a_i) - f(\emptyset) \quad (13)$$
Complement function

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

Proposition

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

Proof.

$$\bar{f}(A) + \bar{f}(B) \geq \bar{f}(A \cup B) + \bar{f}(A \cap B) \quad (14)$$

follows from

$$f(V \setminus A) + f(V \setminus B) \geq f(V \setminus (A \cup B)) + f(V \setminus (A \cap B)) \quad (15)$$

which is true because $V \setminus (A \cup B) = (V \setminus A) \cap (V \setminus B)$ and $V \setminus (A \cap B) = (V \setminus A) \cup (V \setminus B)$.
Submodularity

- Submodular functions have a long history in economics, game theory, combinatorial optimization, electrical networks, and operations research.

- They are gaining importance in machine learning as well (one of our main motivations for offering this course).

- Arbitrary set functions are hopelessly difficult to optimize, while the minimum of submodular functions can be found in polynomial time, and the maximum can be constant-factor approximated in low-order polynomial time.

- Submodular functions share properties in common with both convex and concave functions.
Why do we like Convex Functions? (Quoting Lovász 1983):

1. Convex functions occur in many mathematical models in economy, engineering, and other sciences. Convexity is a very natural property of various functions and domains occurring in such models; quite often the only non-trivial property which can be stated in general.

2. Convexity is preserved under many natural operations and transformations, and thereby the effective range of results can be extended, elegant proof techniques can be developed as well as unforeseen applications of certain results can be given.

3. Convex functions and domains exhibit sufficient structure so that a mathematically beautiful and practically useful theory can be developed.

4. There are theoretically and practically (reasonably) efficient methods to find the minimum of a convex function.
In this course, we wish to demonstrate that submodular functions also possess attractions of these four sorts as well.
Consider an urn containing colored balls. Given a set $S$ of balls, $f(S)$ counts the number of distinct colors.

Submodularity: Incremental Value of Object Diminishes in a Larger Context (diminishing returns).

Thus, $f$ is submodular.
Let $V$ be a set of indices, and each $\nu \in V$ indexes a given sub-area of some region. Let $\text{area}(\nu)$ be the area corresponding to item $\nu$.

Let $f(S) = \bigcup_{s \in S} \text{area}(s)$ be the union of the areas indexed by elements in $A$.

Then $f(S)$ is submodular.
Area of the union of areas indexed by $A$

Union of areas of elements of $A$ is given by:

$$f(A) = f(\{a_1, a_2, a_3, a_4\})$$
Area of the union of areas indexed by $A$

Area of $A$ along with $v$:

$$f(A \cup \{v\}) = f(\{a_1, a_2, a_3, a_4\} \cup \{v\})$$
Gain (value) of $v$ in context of $A$:

$$f(A \cup \{v\}) - f(A) = f(\{v\})$$

We get full value $f(\{v\})$ in this case since the area of $v$ has no overlap with that of $A$. 
Area of the union of areas indexed by $A$

Area of $A$ once again.

$$f(A) = f(\{a_1, a_2, a_3, a_4\})$$
Area of the union of areas indexed by $A$

Union of areas of elements of $B \supset A$, where $v$ is not included:

$$f(B) \text{ where } v \notin B \text{ and where } A \subseteq B$$
Area of the union of areas indexed by $A$

Area of $B$ now also including $v$:

$$f(B \cup \{v\})$$
Incremental value of \( v \) in the context of \( B \).

\[
f(B \cup \{v\}) - f(B) < f(\{v\}) = f(A \cup \{v\}) - f(A)
\]

So benefit of \( v \) in the context of \( A \) is greater than the benefit of \( v \) in the context of \( B \supseteq A \).
Entropy is submodular. Let $V$ be the index set of a set of random variables, then the function

$$f(A) = H(X_A) = - \sum_{x_A} p(x_A) \log p(x_A)$$  \hspace{1cm} (16)

is submodular.

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

$$H(X_v | X_B) = H(X_{B+v}) - H(X_B) \leq H(X_{A+v}) - H(X_A) = H(X_v | X_A)$$  \hspace{1cm} (17)
Example Submodular: Entropy Information Theory

- Alternate Proof: Mutual Information is non-negative.
- Mutual information between two sets of random variables $X_A$ and $X_B$ is given by

$$I(X_A; X_B) = \sum_{X_{A\cup B}} p(x_{A\cup B}) \log \frac{p(x_{A\cup B})p(x_{A\cap B})}{p(x_A)p(x_B)} \geq 0 \quad (18)$$

then

$$I(X_A; X_B) = H(X_A) + H(X_B) - H(X_{A\cup B}) - H(X_{A\cap B}) \geq 0 \quad (19)$$

so entropy satisfies

$$H(X_A) + H(X_B) \geq H(X_{A\cup B}) + H(X_{A\cap B}) \quad (20)$$
Example Submodular: Mutual Information Information Theory

Also, symmetric mutual information is submodular,

\[
f(A) = I(X_A; X_{V\setminus A}) = H(X_A) + H(X_{V\setminus A}) - H(X_V)
\]

(21)

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

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

If \( G \) is undirected, define

\[
E(X, Y) = \{\{x, y\} \in E(G) : x \in X \setminus Y, y \in Y \setminus X\}
\]

as the edges between \( X \) and \( Y \).

Nodes define cuts, and define \( \delta(X) = E(X, V \setminus X) \).

\[
G = (V, E) \quad \Rightarrow \quad S = \{a, b, c\} \quad \delta_G(S) = \{\{u, v\} \in E : u \in S, v \in V \setminus S\} = \{\{a, d\}, \{b, d\}, \{b, e\}, \{c, e\}, \{c, f\}\}
\]
If $G$ is directed, define

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

as the edges from $X$ to $Y$.

Nodes define cuts, and define edges leaving $X$ as $\delta^+(X) = E(X, V \setminus X)$ and edges entering $X$ as $\delta^-(X) = E(V \setminus X, X)$.

$$\delta_G(S) = \{(v, u) \in E : u \in S, v \in V \setminus S\} \quad = \{(d,a) , (d,b), (e,c)\}$$

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

Given a set $X \subseteq V$, the neighbors of $X$ is defined as

$$\Gamma(X) = \{ v \in V(G) \setminus X : E(X, \{v\}) \neq \emptyset \}.$$
Directed Cut functions

Lemma

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

\begin{align}
|\delta^+(X)| + |\delta^+(Y)| &= |\delta^+(X \cap Y)| + |\delta^+(X \cup Y)| + |E^+(X, Y)| + |E^+(Y, X)| \\
|\delta^-(X)| + |\delta^-(Y)| &= |\delta^-(X \cap Y)| + |\delta^-(X \cup Y)| + |E^-(X, Y)| + |E^-(Y, X)|
\end{align}

(22)  
(23)
Proof.

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

We can see the following cases:

1. $\delta^+(X)$
2. $\delta^+(X \cap Y)$
3. $\delta^+(X \cup Y)$
4. $E^+(X, Y)$
5. $E^+(Y, X)$

The cases are illustrated in the diagram.
Directed Cut functions

**Lemma**

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

**Proof.**

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


Lemma

For an undirected graph $G = (V, E)$ and any $X, Y \subseteq V$: we have

\begin{align*}
|\delta(X)| + |\delta(Y)| &= |\delta(X \cap Y)| + |\delta(X \cup Y)| + 2|E(X, Y)| \\
|\Gamma(X)| + |\Gamma(Y)| &\geq |\Gamma(X \cap Y)| + |\Gamma(X \cup Y)|
\end{align*}

Proof.

Eq 24 directly follows from Eq 22 by replacing each edge \{\(u, v\)\} with two oppositely directed edges \((u, v)\) and \((v, u)\) and using the same counting argument.

Eq 25 follows since

\begin{align*}
|\Gamma(X)| + |\Gamma(Y)| &= |\Gamma(X \cup Y)| + |\Gamma(X) \cap \Gamma(Y)| + |\Gamma(X) \cap Y| + |\Gamma(Y) \cap X| \\
&\geq |\Gamma(X \cap Y)| + |\Gamma(X \cup Y)|
\end{align*}
Graphically, we can count and see that

\[ \Gamma(X) = (a) + (c) + (f) + (g) + (d) \]  
\[ \Gamma(Y) = (b) + (c) + (e) + (h) + (d) \]  
\[ \Gamma(X \cup Y) = (a) + (b) + (c) + (d) \]  
\[ \Gamma(X \cap Y) = (c) + (g) + (h) \]

so

\[ |\Gamma(X)| + |\Gamma(Y)| = (a) + (b) + 2(c) + 2(d) + (e) + (f) + (g) + (h) \]  
\[ \geq (a) + (b) + 2(c) + (d) + (g) + (h) = |\Gamma(X \cup Y)| + |\Gamma(X \cap Y)| \]
Therefore, the undirected cut function $\delta(A)$ and the neighbor function $\Gamma(A)$ of a graph $G$ are both submodular.
Other graph functions that are submodular/supermodular

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

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

These come from Narayanan’s book 1997. Let \( G \) be an undirected graph.

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

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

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

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

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

Given a graph, for each $X \subseteq E(G)$, let $c(X)$ denote the number of connected components of the subgraph $(V(G), X)$. Then $c(X)$ is supermodular.

$\bar{c}(X) = C(E \setminus X)$ is the number of connected components in $G$ when we remove $X$, is also supermodular. Maximizing $\bar{c}(X)$ might seem as a goal for a network attacker (choose a small number of edges to sever the graph into as many components as possible).
Matrix Rank functions

Let $V$ be an index set of a set of vectors in $\mathbb{R}^M$ for some $M$. So, for a given set $\{v, v_1, v_2, \ldots, v_k\}$ we might or might not have the possibility of

\[ x_v = \sum_{i=1}^{k} \alpha_i x_i \tag{32} \]

and if not, then $x_v$ is linearly independent of $x_{v_1}, \ldots, x_{v_k}$.

Let $r(S)$ for $S \subseteq V$ be the rank of the set of vectors $S$. Then $r(\cdot)$ is a submodular function, and in fact is called a matric matroid rank function.
Let $S$ be a set of subspaces of a linear space and let, for each $X \subseteq S$, $f(X)$ 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 previous example, and for each $s \in S$, let $X_s$ being an index of vectors. Then, defining

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

is submodular, and is known to be a polymatroid rank function. In general, polymatroid rank function are submodular, normalized $f(\emptyset) = 0$, and 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 edges adjacent to $S$. Then $r(S)$ is submodular, and is another matroid rank function.
A model of Influence in Social Networks

- Given a graph $G = (V, E)$, each $v \in V$ corresponds to a person, to each $v$ we have an activation function $f_v : 2^V \rightarrow [0, 1]$ dependent only on its neighbors. I.e., $f_v(A) = f_v(A \cap \Gamma(v))$.

- Goal: find a small subset $S \subseteq V$ of individuals to directly influence, and thus indirectly influence the greatest number of possible other individuals (via the social network $G$).

- We define a function $f : 2^V \rightarrow \mathbb{Z}^+$ that models the ultimate influence of an initial set $S$ of nodes based on the following iterative process: At each step, a given set of nodes $S$ are activated, and we activate new nodes $v \in V \setminus S$ if $f_v(S) \geq U[0, 1]$ (where $U[0, 1]$ is a uniform random number between 0 and 1).

- It can be shown that for many $f_v$ (including simple linear functions, and where $f_v$ is submodular itself) that $f$ is submodular.
The value of a friend.

- Let $V$ be a group of individuals. How valuable to you is a given friend $v \in V$? It depends on how many friends you have.
- Let $f(S)$ be the value of the set of friends $S$. Is submodular or supermodular a good model?
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.
Economies of scale

So a submodular function would be a good model.
Cost of manufacturing a set of items

Let $V$ be a set of possible items that a company might possibly wish to manufacture, and let $f(S)$ for $S \subseteq V$ be the cost to that company to manufacture subset $S$.

Ex: $V$ might be colors of paint in a paint manufacturer: green, red, blue, yellow, white, etc. Producing green when you are already producing yellow and blue is probably cheaper than if you were only producing some other colors.

$$f(\text{green, blue, yellow}) - f(\text{blue, yellow}) \leq f(\text{green, blue}) - f(\text{blue})$$

(34)

So a submodular function would be a good model.
Submodular Polyhedra

- 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.

\[ P_f = \left\{ x \in \mathbb{R}^E : x \geq 0, x(S) \leq f(S), \forall S \subseteq E \right\} \] (35)

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

\[ \tilde{f}(c) = \max c^T x \text{ such that } x \in P_f \] (36)

- This can be solved using the greedy algorithm! Moreover, \( \tilde{f}(c) \) computed using greedy is convex if and only of \( f \) is submodular (we will go into this in some detail this quarter).
Submodular Definitions

Definition (submodular)

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

\[
f(A) + f(B) \geq f(A \cup B) + f(A \cap B)
\] (1)

An alternate and equivalent definition is:

Definition (diminishing returns)

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

\[
f(A \cup \{v\}) - f(A) \geq f(B \cup \{v\}) - f(B)
\] (2)

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 \( B \).
An alternate and equivalent definition is:

**Definition (group diminishing returns)**

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

\[
f(A \cup C) - f(A) \geq f(B \cup C) - f(B)
\]  

(37)

This means that the incremental “value” or “gain” of set \( C \) decreases as the context in which \( v \) is considered grows from \( A \) to \( B \) (diminishing returns)
Submodular Definitions

Proposition

group diminishing returns implies diminishing returns

Proof.

Obvious, set $C = \{v\}$.

Proposition

diminishing returns implies group diminishing returns
Submodular Definitions (cont. II)

Proof.

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

\[
f(A \cup C) - f(A) = f(A \cup C) - \sum_{i=1}^{k-1} \left( f(A \cup \{c_1, \ldots, c_i\}) - f(A \cup \{c_1, \ldots, c_i\}) \right) - f(A)
\]

\[
= \sum_{i=1}^{k} f(A \cup \{c_1 \ldots c_i\}) - f(A \cup \{c_1 \ldots c_{i-1}\})
\]

\[
\leq \sum_{i=1}^{k} f(B \cup \{c_1 \ldots c_i\}) - f(B \cup \{c_1 \ldots c_{i-1}\})
\]

\[
= f(B \cup C) - \sum_{i=1}^{k-1} \left( f(B \cup \{c_1, \ldots, c_i\}) - f(B \cup \{c_1, \ldots, c_i\}) \right) - f(B)
\]

\[
= f(B \cup C) - f(B)
\]
Submodular Definitions are equivalent

Proposition

The two aforementioned definitions of submodularity submodular and diminishing returns are identical.
Submodular Definitions are equivalent (cont. II)

Proof.

Assume *submodular*. Assume $A \subset B$ as otherwise trivial.

Let $B \setminus A = \{v_1, v_2, \ldots, v_k\}$ and define $A^i = A \cup \{v_1 \ldots v_i\}$, so $A^0 = A$.

Then by *submodular*,

$$f(A^i + v) + f(A^i + v_{i+1}) \geq f(A^i + v + v_{i+1}) + f(A^i) \quad (44)$$

or

$$f(A^i + v) - f(A^i) \geq f(A^i + v_{i+1} + v) - f(A^i + v_{i+1}) \quad (45)$$

we apply this inductively, and use

$$f(A^{i+1} + v) - f(A^{i+1}) = f(A^i + v_{i+1} + v) - f(A^i + v_{i+1}) \quad (46)$$

and that $A^{k-1} + v_k = B$. 

...cont.

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

$$f(A' + C) - f(A') \geq f(B' + C) - f(B') \quad (47)$$

giving

$$f(A' + C) + f(B') \geq f(B' + C) + f(A') \quad (48)$$

or

$$f(A \cap B + A \setminus B) + f(B) \geq f(B + A \setminus B) + f(A \cap B) \quad (49)$$