Submodular Functions, Optimization, and Applications to Machine Learning
— Spring Quarter, Lecture 12 —

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

Prof. Jeff Bilmes

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

May 11th, 2016

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

- \[ f(A) + 2f(C) + f(B) - f(A) + f(C) + f(B) - f(A \cap B) \]
Cumulative Outstanding Reading

- Read chapters 2 and 3 from Fujishige’s book.
- Read chapter 1 from Fujishige’s book.
Homework 4, soon available at our assignment dropbox (https://canvas.uw.edu/courses/1039754/assignments)

Homework 3, available at our assignment dropbox (https://canvas.uw.edu/courses/1039754/assignments), due (electronically) Monday (5/2) at 11:55pm.

Homework 2, available at our assignment dropbox (https://canvas.uw.edu/courses/1039754/assignments), due (electronically) Monday (4/18) at 11:55pm.

Homework 1, available at our assignment dropbox (https://canvas.uw.edu/courses/1039754/assignments), due (electronically) Friday (4/8) at 11:55pm.

Weekly Office Hours: Mondays, 3:30-4:30, or by skype or google hangout (set up meeting via our our discussion board (https://canvas.uw.edu/courses/1039754/discussion_topics)).
Class Road Map - IT-I

L1(3/28): Motivation, Applications, & Basic Definitions
L3(4/4): Info theory exs, more apps, definitions, graph/combinatorial examples, matrix rank example, visualization
L4(4/6): Graph and Combinatorial Examples, matrix rank, Venn diagrams, examples of proofs of submodularity, some useful properties
L5(4/11): Examples & Properties, Other Defs., Independence
L6(4/13): Independence, Matroids, Matroid Examples, matroid rank is submodular
L7(4/18): Matroid Rank, More on Partition Matroid, System of Distinct Reps, Transversals, Transversal Matroid,
L8(4/20): Transversals, Matroid and representation, Dual Matroids,
L10(4/27): Matroid and Greedy, Polyhedra, Matroid Polytopes,
L11(5/2): From Matroids to Polymatroids, Polymatroids
L12(5/4): Polymatroids, Polymatroids and Greedy
L13(5/9): Polymatroids and Greedy; Possible Polytopes; Extreme Points; Polymatroids, Greedy, and Cardinality Constrained Maximization
L14(5/11):
L15(5/16):
L16(5/18):
L17(5/23):
L18(5/25):
L19(6/1):
L20(6/6): Final Presentations maximization.

Finals Week: June 6th-10th, 2016.
Matroid and Polymatroid: side-by-side

A Matroid is:

1. a set system $\mathbf{(E, I)}$
2. empty-set containing $\emptyset \in \mathcal{I}$
3. 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 $P \subseteq \mathbb{R}^E_+$
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 $y(E)$).
A polymatroid vs. a polymatroid function’s polyhedron

- Summarizing the above, we have:
  - Given a polymatroid function $f$, its associated polytope is given as
    \[ P^+_f = \{ y \in \mathbb{R}^E_+ : y(A) \leq f(A) \text{ for all } A \subseteq E \} \] (12.10)
  - We also have the definition of a polymatroidal polytope $P$ (compact subset, zero containing, down-monotone, and $\forall x$ any maximal independent subvector $y \leq x$ has same component sum $y(E)$).

- Is there any relationship between these two polytopes?
- In the next theorem, we show that any $P^+_f$-basis has the same component sum, when $f$ is a polymatroid function, and $P^+_f$ satisfies the other properties so that $P^+_f$ is a polymatroid.
Theorem 12.2.1

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

$$y^x(E) = \text{rank}(x) = \max \left( y(E) : y \leq x, y \in P_f^+ \right)$$

$$= \min \left( x(A) + f(E \setminus A) : A \subseteq E \right)$$

(12.10)

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

Taking $E \setminus B = \text{supp}(x)$ (so elements $B$ are all zeros in $x$), and for $b \notin B$ we make $x(b)$ is big enough, the r.h.s. min has solution $A^* = B$. We recover submodular function from the polymatroid polyhedron via the following:

$$\text{rank} \left( \frac{1}{\epsilon} 1_{E \setminus B} \right) = f(B) = \max \left\{ y(B) : y \in P_f^+ \right\}$$

(12.11)

In fact, we will ultimately see a number of important consequences of this theorem (other than just that $P_f^+$ is a polymatroid)
So, 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 12.2.1**

For any polymatroid $P$ (compact subset of $\mathbb{R}^+_E$, zero containing, down-monotone, and $\forall x \in \mathbb{R}^+_E$ any maximal independent subvector $y \leq x$ has same component sum $y(E) = \text{rank}(x)$), there is a polymatroid function $f : 2^E \to \mathbb{R}$ (normalized, monotone non-decreasing, submodular) such that $P = P^+_f$ where $P^+_f = \{ x \in \mathbb{R}^E : x \geq 0, x(A) \leq f(A), \forall A \subseteq E \}$. 
Tight sets $\mathcal{D}(y)$ are closed, and max tight set $\text{sat}(y)$

Recall the definition of the set of tight sets at $y \in P_f^+$:

$$\mathcal{D}(y) \triangleq \{ A : A \subseteq E, y(A) = f(A) \}$$

(12.10)

Theorem 12.2.1

For any $y \in P_f^+$, with $f$ a polymatroid function, then $\mathcal{D}(y)$ is closed under union and intersection.

Proof.

We have already proven this as part of Theorem 11.4.1

Also recall the definition of $\text{sat}(y)$, the maximal set of tight elements relative to $y \in \mathbb{R}_+^E$.

$$\text{sat}(y) \overset{\text{def}}{=} \bigcup \{ T : T \in \mathcal{D}(y) \}$$

(12.11)
Vector rank, \( \text{rank}(x) \), is submodular

- Recall that the matroid rank function is submodular.
- The vector rank function \( \text{rank}(x) \) also satisfies a form of submodularity, namely one defined on the real lattice.

**Theorem 12.2.1 (vector rank and submodularity)**

Let \( P \) be a polymatroid polytope. The vector rank function \( \text{rank} : \mathbb{R}^E_+ \rightarrow \mathbb{R} \) with \( \text{rank}(x) = \max \{ y(E) : y \leq x, y \in P \} \) satisfies, for all \( u, v \in \mathbb{R}^E_+ \):

\[
\text{rank}(u) + \text{rank}(v) \geq \text{rank}(u \lor v) + \text{rank}(u \land v) \tag{12.10}
\]
Polymatroidal polyhedron and greedy

- Let \((E, \mathcal{I})\) be a set system and \(w \in \mathbb{R}^E_+\) be a weight vector.
- Recall greedy algorithm: Set \(A = \emptyset\), and repeatedly choose \(y \in E \setminus A\) such that \(A \cup \{y\} \in \mathcal{I}\) with \(w(y)\) as large as possible, stopping when no such \(y\) exists.
- For a matroid, we saw that independence system \((E, \mathcal{I})\) is a matroid iff for each weight function \(w \in \mathbb{R}^E_+\), the greedy algorithm leads to a set \(I \in \mathcal{I}\) of maximum weight \(w(I)\).
- Stated succinctly, considering \(\max \{w(I) : I \in \mathcal{I}\}\), then \((E, \mathcal{I})\) is a matroid iff greedy works for this maximization.
- 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?
- Can we, ultimately, even relax things so that \(w \in \mathbb{R}^E\)?
Polymatroidal polyhedron and greedy

- What is the greedy solution in this setting, when \( w \in \mathbb{R}^E \)?
- Sort elements of \( E \) w.r.t. \( w \) so that, w.l.o.g. 
  \[ E = (e_1, e_2, \ldots, e_m) \text{ with } w(e_1) \geq w(e_2) \geq \cdots \geq w(e_m). \]
- Let \( k + 1 \) be the first point (if any) at which we are non-positive, i.e., 
  \( w(e_k) > 0 \) and \( 0 \geq w(e_{k+1}) \).
- Next define partial accumulated sets \( E_i \), for \( i = 0 \ldots m \), we have w.r.t. the above sorted order:
  \[ E_i \overset{\text{def}}{=} \{ e_1, e_2, \ldots, e_i \} \]  
  \( (\text{note } E_0 = \emptyset, \ f(E_0) = 0, \ \text{and } E \text{ and } E_i \text{ is always sorted w.r.t } w) \).
- The greedy solution is the vector \( x \in \mathbb{R}_+^E \) with elements defined as:
  \[ x(e_1) \overset{\text{def}}{=} f(E_1) = f(e_1) = f(e_1|E_0) = f(e_1|\emptyset) \]  
  \[ x(e_i) \overset{\text{def}}{=} f(E_i) - f(E_{i-1}) = f(e_i|E_{i-1}) \text{ for } i = 2 \ldots k \]  
  \[ x(e_i) \overset{\text{def}}{=} 0 \text{ for } i = k + 1 \ldots m = |E| \]
Polymatroidal polyhedron and greedy

**Theorem 12.2.2**

The vector $x \in \mathbb{R}^E_+$ as previously defined using the greedy algorithm maximizes $wx$ over $P_f^+$, with $w \in \mathbb{R}^E_+$, if $f$ is submodular.

**Proof.**

- Consider the LP strong duality equation:

$$\max (wx : x \in P_f^+) = \min \left( \sum_{A \subseteq E} y_A f(A) : y \in \mathbb{R}^{2E}_+, \sum_{A \subseteq E} y_A 1_A \geq w \right)$$

(12.30)

- Sort $E$ by $w$ descending, and define the following vector $y \in \mathbb{R}^{2E}_+$ as

$$y_{E_i} \leftarrow w(e_i) - w(e_{i+1}) \text{ for } i = 1 \ldots (m - 1),$$

(12.31)

$$y_E \leftarrow w(e_m), \text{ and}$$

(12.32)

$$y_A \leftarrow 0 \text{ otherwise}$$

(12.33)
Theorem 12.3.1

Conversely, suppose $P_f^+$ is a polytope of form

$$P_f^+ = \{ x \in \mathbb{R}_+^E : x(A) \leq f(A), \forall A \subseteq E \}$$

then the greedy solution to $\max(wx : x \in P)$ is optimum only if $f$ is submodular.

Proof.

Choose $A$ and $B$ arbitrarily, and then order elements of $E$ as $(e_1, e_2, \ldots, e_m)$, with $E_i = (e_1, e_2, \ldots, e_i)$, so the following is true:
Polymatroidal polyhedron and greedy

**Theorem 12.3.1**

Conversely, suppose $P_f^+$ is a polytope of form

$$P_f^+ = \{ x \in \mathbb{R}_+^E : x(A) \leq f(A), \forall A \subseteq E \},$$

then the greedy solution to $\max(wx : x \in P)$ is optimum only if $f$ is submodular.

**Proof.**

- Choose $A$ and $B$ arbitrarily, and then order elements of $E$ as $(e_1, e_2, \ldots, e_m)$, with $E_i = (e_1, e_2, \ldots, e_i)$, so the following is true:
- For $1 \leq p \leq q \leq m$, $A = \{e_1, e_2, \ldots, e_k, e_{k+1}, \ldots, e_p\} = E_p$ and $B = \{e_1, e_2, \ldots, e_k, e_{p+1}, \ldots, e_q\} = E_k \cup (E_q \setminus E_p) = (A \cup B) \cup (B \setminus A)$
Theorem 12.3.1

Conversely, suppose \( P_f^+ \) is a polytope of form

\[
P_f^+ = \{ x \in \mathbb{R}^E_+ : x(A) \leq f(A), \forall A \subseteq E \},
\]

then the greedy solution to \( \max(wx : x \in P) \) is optimum only if \( f \) is submodular.

Proof.

- Choose \( A \) and \( B \) arbitrarily, and then order elements of \( E \) as \((e_1, e_2, \ldots, e_m)\), with \( E_i = (e_1, e_2, \ldots, e_i) \), so the following is true:

- For \( 1 \leq p \leq q \leq m \), \( A = \{e_1, e_2, \ldots, e_k, e_{k+1}, \ldots, e_p\} = E_p \) and \( B = \{e_1, e_2, \ldots, e_k, e_{p+1}, \ldots, e_q\} = E_k \cup (E_q \setminus E_p) \)

- Note, then we have \( A \cap B = \{e_1, \ldots, e_k\} = E_k \), and \( A \cup B = E_q \).
Theorem 12.3.1

Conversely, suppose $P_f^+$ is a polytope of form

$$P_f^+ = \{ x \in \mathbb{R}_+^E : x(A) \leq f(A), \forall A \subseteq E \},$$

then the greedy solution to

$$\max(wx : x \in P)$$

is optimum only if $f$ is submodular.

Proof.

- Choose $A$ and $B$ arbitrarily, and then order elements of $E$ as $(e_1, e_2, \ldots, e_m)$, with $E_i = (e_1, e_2, \ldots, e_i)$, so the following is true:

- For $1 \leq p \leq q \leq m$, $A = \{e_1, e_2, \ldots, e_k, e_{k+1}, \ldots, e_p\} = E_p$ and $B = \{e_1, e_2, \ldots, e_k, e_{p+1}, \ldots, e_q\} = E_k \cup (E_q \setminus E_p)$

- Note, then we have $A \cap B = \{e_1, \ldots, e_k\} = E_k$, and $A \cup B = E_q$.

- Define $w \in \{0, 1\}^m$ as:

$$w \overset{\text{def}}{=} \sum_{i=1}^{q} 1_{e_i} = 1_{A \cup B} \quad (12.1)$$
Polymatroids and Greedy

**Theorem 12.3.1**

Conversely, suppose $P_f^+$ is a polytope of form

$P_f^+ = \{ x \in \mathbb{R}_+^E : x(A) \leq f(A), \forall A \subseteq E \}$,

then the greedy solution to $\max(wx : x \in P)$ is optimum only if $f$ is submodular.

**Proof.**

- Choose $A$ and $B$ arbitrarily, and then order elements of $E$ as $(e_1, e_2, \ldots, e_m)$, with $E_i = (e_1, e_2, \ldots, e_i)$, so the following is true:
  - For $1 \leq p \leq q \leq m$, $A = \{e_1, e_2, \ldots, e_k, e_{k+1}, \ldots, e_p\} = E_p$ and $B = \{e_1, e_2, \ldots, e_k, e_{p+1}, \ldots, e_q\} = E_k \cup (E_q \setminus E_p)$
  - Note, then we have $A \cap B = \{e_1, \ldots, e_k\} = E_k$, and $A \cup B = E_q$.
- Define $w \in \{0, 1\}^m$ as:
  
  $$w \overset{\text{def}}{=} \sum_{i=1}^{q} 1_{e_i} = 1_{A \cup B}$$  

  (12.1)

- Suppose optimum solution $x$ is given by the greedy procedure.
Proof.

Then

\[
\sum_{i=1}^{k} x_i = f(E_1) + \sum_{i=2}^{k} (f(E_i) - f(E_{i-1})) = f(E_k) = f(A \cap B)
\]

(12.2)
Polymatroidal polyhedron and greedy

**Proof.**

- Then

\[
\sum_{i=1}^{k} x_i = f(E_1) + \sum_{i=2}^{k} (f(E_i) - f(E_{i-1})) = f(E_k) = f(A \cap B) \tag{12.2}
\]

- and

\[
\sum_{i=1}^{p} x_i = f(E_1) + \sum_{i=2}^{p} (f(E_i) - f(E_{i-1})) = f(E_p) = f(A) \tag{12.3}
\]
Proof.

Then

$$\sum_{i=1}^{k} x_i = f(E_1) + \sum_{i=2}^{k} (f(E_i) - f(E_{i-1})) = f(E_k) = f(A \cap B)$$

(12.2)

and

$$\sum_{i=1}^{p} x_i = f(E_1) + \sum_{i=2}^{p} (f(E_i) - f(E_{i-1})) = f(E_p) = f(A)$$

(12.3)

and

$$\sum_{i=1}^{q} x_i = f(E_1) + \sum_{i=2}^{q} (f(E_i) - f(E_{i-1})) = f(E_q) = f(A \cup B)$$

(12.4)
Proof.

Thus, we have

\[ x(B) = \sum_{i=1,\ldots,k,p+1,\ldots,q} x_i = \sum_{i: e_i \in B} x_i = f(A \cup B) + f(A \cap B) - f(A) \]

(12.5)
Polymatroidal polyhedron and greedy

Proof.

- Thus, we have

\[
x(B) = \sum_{i=1,...,k,p+1,...,q} x_i = \sum_{i:e_i \in B} x_i = f(A \cup B) + f(A \cap B) - f(A)
\]

(12.5)

- But given that the greedy algorithm gives the optimal solution to

\[
\max(wx : x \in P_f^+),
\]

we have that \(x \in P_f^+\) and thus \(x(B) \leq f(B)\).
Proof.

Thus, we have

$$x(B) = \sum_{i=1,\ldots,k,p+1,\ldots,q} x_i = \sum_{i:e_i \in B} x_i = f(A \cup B) + f(A \cap B) - f(A)$$

$$= x(A)$$ \quad (12.5)

But given that the greedy algorithm gives the optimal solution to

$$\max(wx : x \in P_f^+)$$

we have that $$x \in P_f^+$$ and thus $$x(B) \leq f(B)$$.

Thus,

$$x(B) = f(A \cup B) + f(A \cap B) - f(A) = \sum_{i:e_i \in B} x_i \leq f(B)$$ \quad (12.6)

ensuring the submodularity of $$f$$, since $$A$$ and $$B$$ are arbitrary.
The next slide comes from lecture 9.
Matroid and the greedy algorithm

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

**Algorithm 1:** The Matroid Greedy Algorithm

1. Set \(X \leftarrow \emptyset\);
2. while \(\exists v \in E \setminus X\) s.t. \(X \cup \{v\} \in I\) do
3. \(v \in \text{argmax} \{w(v) : v \in E \setminus X, X \cup \{v\} \in I\}\);
4. \(X \leftarrow X \cup \{v\}\);

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

**Theorem 12.3.7**

Let \((E, I)\) be an independence system. Then the pair \((E, I)\) is a matroid if and only if for each weight function \(w \in \mathcal{R}_+^E\), Algorithm ?? above leads to a set \(I \in I\) of maximum weight \(w(I)\).
Thus, restating the above results into a single complete theorem, we have a result very similar to what we saw for matroids (i.e., Theorem 10.5.1)

**Theorem 12.3.1**

If $f : 2^E \to \mathbb{R}_+$ is given, and $P$ is a polytope in $\mathbb{R}_+^E$ of the form $P = \{x \in \mathbb{R}_+^E : x(A) \leq f(A), \forall A \subseteq E\}$, then the greedy solution to the problem $\max (wx : x \in P)$ is $\forall w$ optimum iff $f$ is monotone non-decreasing submodular (i.e., iff $P$ is a polymatroid).
Multiple Polytopes associated with arbitrary $f$

- Given an arbitrary submodular function $f : 2^V \rightarrow R$ (not necessarily a polymatroid function, so it need not be positive, monotone, etc.).
Multiple Polytopes associated with arbitrary \( f \)

- Given an arbitrary submodular function \( f : 2^V \rightarrow \mathbb{R} \) (not necessarily a polymatroid function, so it need not be positive, monotone, etc.).

- If \( f(\emptyset) \neq 0 \), can set \( f'(A) = f(A) - f(\emptyset) \) without destroying submodularity. This does not change any minima, (i.e., \( \arg\min_A f(A) = \arg\min_A f'(A) \)) so assume all functions are normalized \( f(\emptyset) = 0 \).

Note that due to constraint \( x(\emptyset) \leq f(\emptyset) \), we must have \( f(\emptyset) \geq 0 \) since if not (i.e., if \( f(\emptyset) < 0 \)), then \( P_f^+ \) doesn’t exist.

Another form of normalization can do is:

\[
f'(A) = \begin{cases} f(A) & \text{if } A \neq \emptyset \\ 0 & \text{if } A = \emptyset \end{cases}
\]

This preserves submodularity due to \( f(A) + f(B) \geq f(A \cup B) + f(A \cap B) \), and if \( A \cap B = \emptyset \) then r.h.s. only gets smaller when \( f(\emptyset) \geq 0 \).
Multiple Polytopes associated with arbitrary $f$

- Given an arbitrary submodular function $f : 2^V \rightarrow \mathbb{R}$ (not necessarily a polymatroid function, so it need not be positive, monotone, etc.).
- If $f(\emptyset) \neq 0$, can set $f'(A) = f(A) - f(\emptyset)$ without destroying submodularity. This does not change any minima, (i.e., $\text{argmin}_A f(A) = \text{argmin}_A f'(A)$) so assume all functions are normalized $f(\emptyset) = 0$.

We can define several polytopes:

\begin{align*}
P_f &= \{ x \in \mathbb{R}^E : x(S) \leq f(S), \forall S \subseteq E \} \quad (12.7) \\
P_f^+ &= P_f \cap \{ x \in \mathbb{R}^E : x \geq 0 \} \quad (12.8) \\
B_f &= P_f \cap \{ x \in \mathbb{R}^E : x(E) = f(E) \} \quad (12.9)
\end{align*}
Multiple Polytopes associated with arbitrary $f$

- Given an arbitrary submodular function $f : 2^V \to R$ (not necessarily a polymatroid function, so it need not be positive, monotone, etc.).
- If $f(\emptyset) \neq 0$, can set $f'(A) = f(A) - f(\emptyset)$ without destroying submodularity. This does not change any minima, (i.e., $\text{argmin}_A f(A) = \text{argmin}_A f'(A)$) so assume all functions are normalized $f(\emptyset) = 0$.
- We can define several polytopes:

\begin{align*}
P_f &= \{ x \in \mathbb{R}^E : x(S) \leq f(S), \forall S \subseteq E \} \quad (12.7) \\
P_f^+ &= P_f \cap \{ x \in \mathbb{R}^E : x \geq 0 \} \quad (12.8) \\
B_f &= P_f \cap \{ x \in \mathbb{R}^E : x(E) = f(E) \} \quad (12.9)
\end{align*}

- $P_f$ is sometimes called the extended polytope (sometimes notated as $EP_f$.)
Multiple Polytopes associated with arbitrary $f$

- Given an arbitrary submodular function $f : 2^V \rightarrow \mathbb{R}$ (not necessarily a polymatroid function, so it need not be positive, monotone, etc.).
- If $f(\emptyset) \neq 0$, can set $f'(A) = f(A) - f(\emptyset)$ without destroying submodularity. This does not change any minima, (i.e., $\text{argmin}_A f(A) = \text{argmin}_{A'} f'(A)$) so assume all functions are normalized $f(\emptyset) = 0$.
- We can define several polytopes:

$$P_f = \left\{ x \in \mathbb{R}^E : x(S) \leq f(S), \forall S \subseteq E \right\} \quad (12.7)$$

$$P_f^+ = P_f \cap \left\{ x \in \mathbb{R}^E : x \geq 0 \right\} \quad (12.8)$$

$$B_f = P_f \cap \left\{ x \in \mathbb{R}^E : x(E) = f(E) \right\} \quad (12.9)$$

- $P_f$ is what is sometimes called the extended polytope (sometimes notated as $EP_f$).
- $P_f^+$ is $P_f$ restricted to the positive orthant.
Given an arbitrary submodular function $f : 2^V \to \mathbb{R}$ (not necessarily a polymatroid function, so it need not be positive, monotone, etc.).

If $f(\emptyset) \neq 0$, can set $f'(A) = f(A) - f(\emptyset)$ without destroying submodularity. This does not change any minima, (i.e., $\text{argmin}_A f(A) = \text{argmin}_{A'} f'(A)$) so assume all functions are normalized $f(\emptyset) = 0$.

We can define several polytopes:

\begin{align*}
P_f &= \{x \in \mathbb{R}^E : x(S) \leq f(S), \forall S \subseteq E\} \\
P_f^+ &= P_f \cap \{x \in \mathbb{R}^E : x \geq 0\} \\
B_f &= P_f \cap \{x \in \mathbb{R}^E : x(E) = f(E)\}
\end{align*}

$P_f$ is what is sometimes called the extended polytope (sometimes notated as $EP_f$).

$P_f^+$ is $P_f$ restricted to the positive orthant.

$B_f$ is called the base polytope, analogous to the base in matroid.
Multiple Polytopes associated with $f$ in $\mathcal{F}$.

$P^+_f = P_f \cap \{x \in \mathbb{R}^E : x \geq 0\} \quad (12.10)$

$P_f = \{x \in \mathbb{R}^E : x(S) \leq f(S), \forall S \subseteq E\} \quad (12.11)$

$B_f = P_f \cap \{x \in \mathbb{R}^E : x(E) = f(E)\} \quad (12.12)$
Multiple Polytopes associated with $f$

\[ P_f^+ = P_f \cap \{ x \in \mathbb{R}^E : x \geq 0 \} \]  
\[ P_f = \{ x \in \mathbb{R}^E : x(S) \leq f(S), \forall S \subseteq E \} \]  
\[ B_f = P_f \cap \{ x \in \mathbb{R}^E : x(E) = f(E) \} \]  
\[ (12.10) \]
\[ (12.11) \]
\[ (12.12) \]
Multiple Polytopes associated with $f$ 

$P^+_f = P_f \cap \{ x \in \mathbb{R}^E : x \geq 0 \}$ \hspace{1cm} (12.10)

$P_f = \{ x \in \mathbb{R}^E : x(S) \leq f(S), \forall S \subseteq E \}$ \hspace{1cm} (12.11)

$B_f = P_f \cap \{ x \in \mathbb{R}^E : x(E) = f(E) \}$ \hspace{1cm} (12.12)
Polymatroids and Greedy Possible Polytopes Extreme Points Polymatroids, Greedy, and Cardinality Constrained Maximization

**Base Polytope in 3D**

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

---

**Equations**

(12.13)

(12.14)
A polymatroid function’s polyhedron is a polymatroid.

**Theorem 12.4.1**

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

\[
\text{rank}(x) = \max \{ y(E) : y \leq x, y \in P_f \} = \min \{ x(A) + f(E \setminus A) : A \subseteq E \} \tag{12.15}
\]

Essentially the same theorem as Theorem 11.4.1, but note \( P_f \) rather than \( P_f^+ \). Taking \( x = 0 \) we get:

**Corollary 12.4.2**

Let \( f \) be a submodular function defined on subsets of \( E \). We have:

\[
\text{rank}(0) = \max \{ y(E) : y \leq 0, y \in P_f \} = \min \{ f(A) : A \subseteq E \} \tag{12.16}
\]
Proof of Theorem 12.4.1.

Let $y^*$ be the optimal solution of the l.h.s. and let $A \subseteq E$ be any subset.
Proof of Theorem 12.4.1

Let $y^*$ be the optimal solution of the l.h.s. and let $A \subseteq E$ be any subset.

Then $y^*(E) = y^*(A) + y^*(E \setminus A) \leq f(A) + x(E \setminus A)$ since if $y^* \in P_f$, $y^*(A) \leq f(A)$ and since $y^* \leq x$, $y^*(E \setminus A) \leq x(E \setminus A)$. This is a form of weak duality.
Proof of Theorem 12.4.1

Let $y^*$ be the optimal solution of the l.h.s. and let $A \subseteq E$ be any subset.

Then $y^*(E) = y^*(A) + y^*(E \setminus A) \leq f(A) + x(E \setminus A)$ since if $y^* \in P_f$, $y^*(A) \leq f(A)$ and since $y^* \leq x$, $y^*(E \setminus A) \leq x(E \setminus A)$. This is a form of weak duality.

Also, for any $e \in E$, if $y^*(e) < x(e)$ then there must be some reason for this other than the constraint $y^* \leq x$, namely it must be that $\exists T \in \mathcal{D}(x)$ with $e \in T$ (i.e., $e$ is a member of at least one of the tight sets).
Proof of Theorem 12.4.1.

Let \( y^* \) be the optimal solution of the l.h.s. and let \( A \subseteq E \) be any subset.

Then \( y^*(E) = y^*(A) + y^*(E \setminus A) \leq f(A) + x(E \setminus A) \) since if \( y^* \in P_f \), \( y^*(A) \leq f(A) \) and since \( y^* \leq x \), \( y^*(E \setminus A) \leq x(E \setminus A) \). This is a form of weak duality.

Also, for any \( e \in E \), if \( y^*(e) < x(e) \) then there must be some reason for this other than the constraint \( y^* \leq x \), namely it must be that \( \exists T \in \mathcal{D}(x) \) with \( e \in T \) (i.e., \( e \) is a member of at least one of the tight sets).

Hence, for all \( e \notin \text{sat}(y^*) \) we have \( y^*(e) = x(e) \), and moreover \( y^*(\text{sat}(y^*)) = f(\text{sat}(y^*)) \) by definition.
Proof of Theorem 12.4.1.

Let $y^*$ be the optimal solution of the l.h.s. and let $A \subseteq E$ be any subset.

Then $y^*(E) = y^*(A) + y^*(E \setminus A) \leq f(A) + x(E \setminus A)$ since if $y^* \in Pf$, $y^*(A) \leq f(A)$ and since $y^* \leq x$, $y^*(E \setminus A) \leq x(E \setminus A)$. This is a form of weak duality.

Also, for any $e \in E$, if $y^*(e) < x(e)$ then there must be some reason for this other than the constraint $y^* \leq x$, namely it must be that $\exists T \in \mathcal{D}(x)$ with $e \in T$ (i.e., $e$ is a member of at least one of the tight sets).

Hence, for all $e \notin \text{sat}(y^*)$ we have $y^*(e) = x(e)$, and moreover $y^*(\text{sat}(y^*)) = f(\text{sat}(y^*))$ by definition.

Thus we have that $y^*(\text{sat}(y^*)) + y^*(E \setminus \text{sat}(y^*)) = f(\text{sat}(y^*)) + x(E \setminus \text{sat}(y^*))$, strong duality, showing that the two sides are equal for $y^*$. 
Greedy and $P_f$

- In Theorem ??, we can relax $P_f^+$ to $P_f$ (prime and dual feasibility still hold as does strong duality).

$$\max_{w, x: x \in P_f^+} \text{subject to greedy.}$$

\[\text{Diagram:}\]

\[\text{Diagram:}\]
Greedy and $P_f$

- In Theorem ??, we can relax $P_f^+$ to $P_f$ (prime and dual feasibility still hold as does strong duality).
- The proof, that is, shows that $x \in P_f$, not just $P_f^+$. 

If $e$ such that $w(e) < 0$, then $\max(\{wx : x \in P_f\}) = 1$ since we can let $x \rightarrow -e$, unless we ignore the negative elements or assume $w \geq 0$. Moreover, in either $P_f$ or $P_f^+$ case, since the greedy constructed $x$ with $x(E) = f(E)$, we have that the greedy $x \in \mathcal{B}_f$. In fact, we will see, in the next section, that the greedy $x$ is a vertex of $\mathcal{B}_f$. 

Prof. Jeff Bilmes
In Theorem ??, we can relax $P_f^+$ to $P_f$ (prime and dual feasibility still hold as does strong duality).

The proof, that is, shows that $x \in P_f$, not just $P_f^+$.

If $\exists e$ such that $w(e) < 0$ then $\max(wx : x \in P_f) = \infty$ since we can let $x_e \to \infty$, unless we ignore the negative elements or assume $w \geq 0$. 

$\infty \left( x \in B_f \right)$. 
Greedy and $P_f$

- In Theorem ??, we can relax $P_f^+$ to $P_f$ (prime and dual feasibility still hold as does strong duality).
- The proof, that is, shows that $x \in P_f$, not just $P_f^+$.
- If $\exists e$ such that $w(e) < 0$ then $\max(wx : x \in P_f) = \infty$ since we can let $x_e \rightarrow \infty$, unless we ignore the negative elements or assume $w \geq 0$.
- Moreover, in either $P_f$, or $P_f^+$ case, since the greedy constructed an $x$ has $x(E) = f(E)$, we have that the greedy $x \in B_f$.

\[
\begin{align*}
  x(e_i) &= f(e_i) \\
  x(e_i e_j) &= f(e_i) - f(e_j) = f(e_i) \\
  x(e_i \setminus e_j) &= f(e_i / e_j) \\
  \sum x(e_i) &= f(E)
\end{align*}
\]
In Theorem ??, we can relax $P_f^+$ to $P_f$ (prime and dual feasibility still hold as does strong duality).

The proof, that is, shows that $x \in P_f$, not just $P_f^+$.

If $\exists e$ such that $w(e) < 0$ then $\max(wx : x \in P_f) = \infty$ since we can let $x_e \to \infty$, unless we ignore the negative elements or assume $w \geq 0$.

Moreover, in either $P_f$, or $P_f^+$ case, since the greedy constructed an $x$ has $x(E) = f(E)$, we have that the greedy $x \in B_f$.

In fact, we will see, in the next section, that the greedy $x$ is a vertex of $B_f$. 
Recall that Theorem 11.4.1 states that
\[
\max \left( y(E) : y \leq x, y \in P_f^+ \right) = \min \left( x(A) + f(E \setminus A) : A \subseteq E \right)
\]
Recall that Theorem 11.4.1 states that
\[
\max \left( y(E) : y \leq x, y \in P_f^+ \right) = \min \left( x(A) + f(E \setminus A) : A \subseteq E \right)
\]

Theorem ?? states that greedy algorithm maximizes \(wx\) over \(P_f^+\) for \(w \in \mathbb{R}_E^+\) with \(f\) being submodular.
Recall that Theorem 11.4.1 states that
\[ \max \left( y(E) : y \leq x, y \in P_f^+ \right) = \min \left( x(A) + f(E \setminus A) : A \subseteq E \right) \]

Theorem ?? states that greedy algorithm maximizes \( wx \) over \( P_f^+ \) for \( w \in \mathbb{R}_E^+ \) with \( f \) being submodular.

Above implies that Theorem ?? can be generalized to \( P_f \) and that greedy solution gives a point in \( B_f \), even for arbitrary finite \( w \).
Polymatroid extreme points

- The greedy algorithm does more than solve $\max(wx : x \in P_f^+)$. We can use it to generate vertices of polymatroidal polytopes.
Polymatroid extreme points

- The greedy algorithm does more than solve $\max(wx : x \in P_f^+)$.
  We can use it to generate vertices of polymatroidal polytopes.
- Consider $P_f^+$ and also $C_f^+ \overset{\text{def}}{=} \{x : x \in \mathbb{R}_+^E, x(e) \leq f(e), \forall e \in E\}$

$P_f^+ \subseteq C_f^+$
Polymatroid extreme points

- The greedy algorithm does more than solve $\max(wx : x \in P_f^+)$. We can use it to generate vertices of polymatroidal polytopes.
- Consider $P_f^+$ and also $C_f^+ \overset{\text{def}}{=} \{ x : x \in \mathbb{R}_+^E, x(e) \leq f(e), \forall e \in E \}$
- Then ordering $A = (a_1, \ldots, a_{|A|})$ arbitrarily with $A_i = \{a_1, \ldots, a_i\}$, $f(A) = \sum_i f(a_i|A_{i-1}) \leq \sum_i f(a_i)$, and hence $P_f^+ \subseteq C_f^+$. 
The greedy algorithm does more than solve \( \max(wx : x \in P_f^+) \). We can use it to generate vertices of polymatroidal polytopes.

Consider \( P_f^+ \) and also \( C_f^+ \) defined as \( \{ x : x \in \mathbb{R}_+^E, x(e) \leq f(e), \forall e \in E \} \).

Then ordering \( A = (a_1, \ldots, a_{|A|}) \) arbitrarily with \( A_i = \{a_1, \ldots, a_i\} \),

\[
 f(A) = \sum_i f(a_i|A_{i-1}) \leq \sum_i f(a_i), \text{ and hence } P_f^+ \subseteq C_f^+. 
\]
Polymatroid extreme points

Since \( w \in \mathbb{R}_+^E \) is arbitrary, it may be that any \( e \in E \) is max (i.e., is such that \( w(e) > w(e') \) for \( e' \in E \setminus \{e\} \)).

\[
\max w \cdot x \quad x \in P_f^+
\]
Polymatroid extreme points

Since \( w \in \mathbb{R}^E_+ \) is arbitrary, it may be that any \( e \in E \) is max (i.e., is such that \( w(e) > w(e') \) for \( e' \in E \setminus \{e\} \)).

Thus, intuitively, any first vertex of the polytope away from the origin might be obtained by advancing along the corresponding axis.

\[
\begin{align*}
\circ & \text{ Suppose } e_1, i \text{ mix, } w(e_1) \geq w(e_i), i \neq 1. \\
\circ & \quad w(e_1), x(e_1) \\
\circ & \quad w(e_2), f(e_2 \mid e_1, \text{ stuff}) \\
& \quad \downarrow \\
& \quad \emptyset
\end{align*}
\]
Polymatroid extreme points

- Since \( w \in \mathbb{R}_+^E \) is arbitrary, it may be that any \( e \in E \) is max (i.e., is such that \( w(e) > w(e') \) for \( e' \in E \setminus \{e\} \)).

- Thus, intuitively, any first vertex of the polytope away from the origin might be obtained by advancing along the corresponding axis.

- Recall, base polytope defined as the extreme face of \( P_f \). I.e.,

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

(12.17)
Polymatroid extreme points

Since \( w \in \mathbb{R}_+^E \) is arbitrary, it may be that any \( e \in E \) is max (i.e., is such that \( w(e) > w(e') \) for \( e' \in E \setminus \{e\} \)).

Thus, intuitively, any first vertex of the polytope away from the origin might be obtained by advancing along the corresponding axis.

Recall, base polytope defined as the extreme face of \( P_f \). I.e.,

\[
B_f = P_f \cap \{ x \in \mathbb{R}_+^E : x(E) = f(E) \} \tag{12.17}
\]
Polymatroid extreme points

- Since \( w \in \mathbb{R}^E_+ \) is arbitrary, it may be that any \( e \in E \) is max (i.e., is such that \( w(e) > w(e') \) for \( e' \in E \setminus \{e\} \)).

- Thus, intuitively, any first vertex of the polytope away from the origin might be obtained by advancing along the corresponding axis.

- Recall, base polytope defined as the extreme face of \( P_f \). I.e.,

\[
B_f = P_f \cap \left\{ x \in \mathbb{R}^E_+ : x(E) = f(E) \right\} \tag{12.17}
\]

- Also, intuitively, we can continue advancing along the skeletal edges of the polytope to reach any other vertex, given the appropriate ordering. If we advance in all dimensions, we’ll reach a vertex in \( B_f \), and if we advance only in some dimensions, we’ll reach a vertex in \( P_f \setminus B_f \).
Polymatroid extreme points

- Since $w \in \mathbb{R}^E_+$ is arbitrary, it may be that any $e \in E$ is max (i.e., is such that $w(e) > w(e')$ for $e' \in E \setminus \{e\}$).

- Thus, intuitively, any first vertex of the polytope away from the origin might be obtained by advancing along the corresponding axis.

- Recall, base polytope defined as the extreme face of $P_f$. I.e.,

$$B_f = P_f \cap \{x \in \mathbb{R}^E_+ : x(E) = f(E)\} \quad (12.17)$$

- Also, intuitively, we can continue advancing along the skeletal edges of the polytope to reach any other vertex, given the appropriate ordering. If we advance in all dimensions, we’ll reach a vertex in $B_f$, and if we advance only in some dimensions, we’ll reach a vertex in $P_f \setminus B_f$.

- We formalize this next:
Polymatroid extreme points

Given any arbitrary order of $E = (e_1, e_2, \ldots, e_m)$, define $E_i = (e_1, e_2, \ldots, e_i)$. 

As before, a vector $x$ is generated by $E_i$ using the greedy procedure as follows:

$$x(e_1) = f(E_1) = f(e_1) \quad (12.18)$$

$$x(e_j) = f(E_j) = f(E_j \cup e_1) \quad \text{for} \ 2 \leq j \leq i \quad (12.19)$$

$$x(e) = 0 \quad \text{for} \ e \in E \cap E_i \quad (12.20)$$
Polymatroid extreme points

- Given any arbitrary order of \( E = (e_1, e_2, \ldots, e_m) \), define \( E_i = (e_1, e_2, \ldots, e_i) \).

- As before, a vector \( x \) is generated by \( E_i \) using the greedy procedure as follows

\[
x(e_1) = f(E_1) = f(e_1)
\]

\[
x(e_j) = f(E_j) - f(E_{j-1}) = f(e_j | E_{j-1}) \quad \text{for } 2 \leq j \leq i
\]

\[
x(e) = 0 \quad \text{for } e \in E \setminus E_i
\]
Polymatroid extreme points

- Given any arbitrary order of $E = (e_1, e_2, \ldots, e_m)$, define $E_i = (e_1, e_2, \ldots, e_i)$.

- As before, a vector $x$ is generated by $E_i$ using the greedy procedure as follows

  \[
  x(e_1) = f(E_1) = f(e_1) \tag{12.18}
  \]

  \[
  x(e_j) = f(E_j) - f(E_{j-1}) = f(e_j|E_{j-1}) \text{ for } 2 \leq j \leq i \tag{12.19}
  \]

  \[
  x(e) = 0 \text{ for } e \in E \setminus E_i \tag{12.20}
  \]

- An extreme point of $P_f$ is a point that is not a convex combination of two other distinct points in $P_f$. Equivalently, an extreme point corresponds to setting certain inequalities in the specification of $P_f$ to be equalities, so that there is a unique single point solution.
Polymatroid extreme points

Theorem 12.5.1

For a given ordering \( E = (e_1, \ldots, e_m) \) of \( E \) and a given \( E_i = (e_1, \ldots, e_i) \) and \( x \) generated by \( E_i \) using the greedy procedure \( (x(e_i) = f(e_i|E_{i-1})) \), then \( x \) is an extreme point of \( P_f \).
Theorem 12.5.1

For a given ordering $E = (e_1, \ldots, e_m)$ of $E$ and a given $E_i = (e_1, \ldots, e_i)$ and $x$ generated by $E_i$ using the greedy procedure ($x(e_i) = f(e_i | E_{i-1})$), then $x$ is an extreme point of $P_f$

Proof.

- We already saw that $x \in P_f$ (Theorem ??).
Polymatroid extreme points

Theorem 12.5.1

For a given ordering \( E = (e_1, \ldots, e_m) \) of \( E \) and a given \( E_i = (e_1, \ldots, e_i) \) and \( x \) generated by \( E_i \) using the greedy procedure \( (x(e_i) = f(e_i | E_{i-1})) \), then \( x \) is an extreme point of \( P_f \).

Proof.

- We already saw that \( x \in P_f \) (Theorem ??).
- To show that \( x \) is an extreme point of \( P_f \), note that it is the unique solution of the following system of equations

\[
\begin{align*}
x(E_j) &= f(E_j) \quad \text{for } 1 \leq j \leq i \leq m \quad (12.21) \\
x(e) &= 0 \quad \text{for } e \in E \setminus E_i \quad (12.22)
\end{align*}
\]

There are \( i \leq m \) equations and \( i \leq m \) unknowns, and simple Gaussian elimination gives us back the \( x \) constructed via the Greedy algorithm!!
Polymatroid extreme points

As an example, we have $x(E_1) = x(e_1) = f(e_1)$
Polymatroid extreme points

As an example, we have $x(E_1) = x(e_1) = f(e_1)$

$x(E_2) = x(e_1) + x(e_2) = f(e_1, e_2)$ so

$x(e_2) = f(e_1, e_2) - x(e_1) = f(e_1, e_2) - f(e_1) = f(e_2 | e_1)$. 
As an example, we have $x(E_1) = x(e_1) = f(e_1)$

$x(E_2) = x(e_1) + x(e_2) = f(e_1, e_2)$ so

$x(e_2) = f(e_1, e_2) - x(e_1) = f(e_1, e_2) - f(e_1) = f(e_2 | e_1)$.

$x(E_3) = x(e_1) + x(e_2) + x(e_3) = f(e_1, e_2, e_3)$ so $x(e_3) = f(e_1, e_2, e_3) - x(e_2) - x(e_1) = f(e_1, e_2, e_3) - f(e_1, e_2) = f(e_3 | e_1, e_2)$
Polymatroid extreme points

- As an example, we have $x(E_1) = x(e_1) = f(e_1)$
- $x(E_2) = x(e_1) + x(e_2) = f(e_1, e_2)$ so $x(e_2) = f(e_1, e_2) - x(e_1) = f(e_1, e_2) - f(e_1) = f(e_2|e_1)$.
- $x(E_3) = x(e_1) + x(e_2) + x(e_3) = f(e_1, e_2, e_3)$ so $x(e_3) = f(e_1, e_2, e_3) - x(e_2) - x(e_1) = f(e_1, e_2, e_3) - f(e_1, e_2) = f(e_3|e_1, e_2)$
- And so on . . . , but we see that this is just Gaussian elimination.
Polymatroid extreme points

- As an example, we have \( x(E_1) = x(e_1) = f(e_1) \)
- \( x(E_2) = x(e_1) + x(e_2) = f(e_1, e_2) \) so
  \[ x(e_2) = f(e_1, e_2) - x(e_1) = f(e_1, e_2) - f(e_1) = f(e_2 | e_1). \]
- \( x(E_3) = x(e_1) + x(e_2) + x(e_3) = f(e_1, e_2, e_3) \) so \( x(e_3) = f(e_1, e_2, e_3) - x(e_2) - x(e_1) = f(e_1, e_2, e_3) - f(e_1, e_2) = f(e_3 | e_1, e_2) \)

- And so on ..., but we see that this is just Gaussian elimination.
- Also, since \( x \in P_f \), for each \( i \), we see that,

\[
\begin{align*}
x(E_j) &= f(E_j) & \text{for } 1 \leq j \leq i \\
x(A) &\leq f(A), \forall A \subseteq E
\end{align*}
\]

Thus, the greedy procedure provides a modular function lower bound on \( f \) that is tight on all points \( E_i \) in the order. This can be useful in its own right, as it provides subgradients and subdi↵erential structure.
As an example, we have \( x(E_1) = x(e_1) = f(e_1) \)
\[
x(E_2) = x(e_1) + x(e_2) = f(e_1, e_2) \quad \text{so} \quad x(e_2) = f(e_1, e_2) - x(e_1) = f(e_1, e_2) - f(e_1) = f(e_2|e_1).
\]
\[
x(E_3) = x(e_1) + x(e_2) + x(e_3) = f(e_1, e_2, e_3) \quad \text{so} \quad x(e_3) = f(e_1, e_2, e_3) - x(e_2) - x(e_1) = f(e_1, e_2, e_3) - f(e_1, e_2) = f(e_3|e_1, e_2)
\]
And so on . . . , but we see that this is just Gaussian elimination.

Also, since \( x \in P_f \), for each \( i \), we see that,
\[
\begin{align*}
x(E_j) &= f(E_j) \quad \text{for } 1 \leq j \leq i \\
x(A) &\leq f(A), \forall A \subseteq E
\end{align*}
\]
Thus, the greedy procedure provides a modular function lower bound on \( f \) that is tight on all points \( E_i \) in the order. This can be useful in its own right, as it provides subgradients and subdifferential structure.
Polymatroid extreme points

some examples
Polymatroid extreme points

Moreover, we have (and will ultimately prove)

**Corollary 12.5.2**

If $x$ is an extreme point of $P_f$ and $B \subseteq E$ is given such that

$\text{supp}(x) = \{e \in E : x(e) \neq 0\} \subseteq B \subseteq \bigcup (A : x(A) = f(A)) = \text{sat}(x)$, then

$x$ is generated using greedy by some ordering of $B$. 

For arbitrary $x$, $\text{supp}(x)$ is not necessarily tight, but for an extreme point, $\text{supp}(x)$ is.
Polymatroid extreme points

Moreover, we have (and will ultimately prove)

**Corollary 12.5.2**

*If \( x \) is an extreme point of \( P_f \) and \( B \subseteq E \) is given such that \( \text{supp}(x) = \{ e \in E : x(e) \neq 0 \} \subseteq B \subseteq \bigcup (A : x(A) = f(A)) = \text{sat}(x) \), then \( x \) is generated using greedy by some ordering of \( B \).*

*Note, \( \text{sat}(x) = \text{cl}(x) = \bigcup (A : x(A) = f(A)) \) is also called the closure of \( x \) (recall that sets \( A \) such that \( x(A) = f(A) \) are called tight, and such sets are closed under union and intersection, as seen in Lecture 8, Theorem ??).*
Polymatroid extreme points

Moreover, we have (and will ultimately prove)

**Corollary 12.5.2**

*If* $x$ *is an extreme point of* $P_f$ *and* $B \subseteq E$ *is given such that*

$$\text{supp}(x) = \{e \in E : x(e) \neq 0\} \subseteq B \subseteq \bigcup(A : x(A) = f(A)) = \text{sat}(x),$$

*then* $x$ *is generated using greedy by some ordering of* $B$.

- Note, $\text{sat}(x) = \text{cl}(x) = \bigcup(A : x(A) = f(A))$ *is also called the closure of* $x$ *(recall that sets* $A$ *such that* $x(A) = f(A)$ *are called tight, and such sets are closed under union and intersection, as seen in Lecture 8, Theorem ??)*

- Thus, $\text{cl}(x)$ *is a tight set.*
Polymatroid extreme points

Moreover, we have (and will ultimately prove)

**Corollary 12.5.2**

*If x is an extreme point of $P_f$ and $B \subseteq E$ is given such that $\text{supp}(x) = \{e \in E : x(e) \neq 0\} \subseteq B \subseteq \bigcup(A : x(A) = f(A)) = \text{sat}(x)$, then $x$ is generated using greedy by some ordering of $B$.***

Note, $\text{sat}(x) = \text{cl}(x) = \bigcup(A : x(A) = f(A))$ is also called the closure of $x$ (recall that sets $A$ such that $x(A) = f(A)$ are called tight, and such sets are closed under union and intersection, as seen in Lecture 8, Theorem ??)

Thus, $\text{cl}(x)$ is a tight set.

Also, $\text{supp}(x) = \{e \in E : x(e) \neq 0\}$ is called the support of $x$. 
Polymatroid extreme points

Moreover, we have (and will ultimately prove)

**Corollary 12.5.2**

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

- Note, $\text{sat}(x) = \text{cl}(x) = \bigcup(A : x(A) = f(A))$ is also called the closure of $x$ (recall that sets $A$ such that $x(A) = f(A)$ are called tight, and such sets are closed under union and intersection, as seen in Lecture 8, Theorem ??)
- Thus, $\text{cl}(x)$ is a tight set.
- Also, supp$(x) = \{e \in E : x(e) \neq 0\}$ is called the support of $x$.
- For arbitrary $x$, supp$(x)$ is not necessarily tight, but for an extreme point, supp$(x)$ is.
Polymatroid with labeled edge lengths

- Recall
  \[ f(e|A) = f(A + e) - f(A) \]

- Notice how submodularity,
  \[ f(e|B) \leq f(e|A) \] for \( A \subseteq B \), defines the shape of the polytope.

- In fact, we have strictness here
  \[ f(e|B) < f(e|A) \] for \( A \subset B \).

- Also, consider how the greedy algorithm proceeds along the edges of the polytope.
Recall 
\[ f(e|A) = f(A+e) - f(A) \]

Notice how submodularity, 
\[ f(e|B) \leq f(e|A) \] for \( A \subseteq B \), defines the shape of the polytope.

In fact, we have strictness here
\[ f(e|B) < f(e|A) \] for \( A \subset B \).

Also, consider how the greedy algorithm proceeds along the edges of the polytope.
Given $w$, the goal is to find
$x = (x(e_1), x(e_2))$
that maximizes
$x^Tw = x(e_1)w(e_1) + x(e_2)w(e_2)$.

If $w(e_2) > w(e_1)$ the upper extreme point indicated maximizes $x^Tw$ over $x \in P^+_f$.

If $w(e_2) < w(e_1)$ the lower extreme point indicated maximizes $x^Tw$ over $x \in P^+_f$. 
Submodular maximization is quite useful.
Maximization of Submodular Functions

- Submodular maximization is quite useful.
- Applications: sensor placement, facility location, document summarization, or any kind of covering problem (choose a small set of elements that cover some domain as much as possible).
Maximization of Submodular Functions

- Submodular maximization is quite useful.
- Applications: sensor placement, facility location, document summarization, or any kind of covering problem (choose a small set of elements that cover some domain as much as possible).
- For polymatroid function (or any monotone non-decreasing function), unconstrained maximization is trivial (take ground set).
Submodular maximization is quite useful.

Applications: sensor placement, facility location, document summarization, or any kind of covering problem (choose a small set of elements that cover some domain as much as possible).

For polymatroid function (or any monotone non-decreasing function), unconstrained maximization is trivial (take ground set).

Thus, when we do monotone submodular maximization, we either

Find the maximum under some constraint
Find the maximum for a non-polymatroid submodular function
Do both.

There is also a sort of dual problem that is often considered together with max, and those are minimum cover problems (to be defined).
Submodular maximization is quite useful.

Applications: sensor placement, facility location, document summarization, or any kind of covering problem (choose a small set of elements that cover some domain as much as possible).

For polymatroid function (or any monotone non-decreasing function), unconstrained maximization is trivial (take ground set).

Thus, when we do monotone submodular maximization, we either

- Find the maximum under some constraint
Submodular maximization is quite useful.

Applications: sensor placement, facility location, document summarization, or any kind of covering problem (choose a small set of elements that cover some domain as much as possible).

For polymatroid function (or any monotone non-decreasing function), unconstrained maximization is trivial (take ground set).

Thus, when we do monotone submodular maximization, we either
- Find the maximum under some constraint
- Find the maximum for a non-polymatroid submodular function
Submodular maximization is quite useful.

Applications: sensor placement, facility location, document summarization, or any kind of covering problem (choose a small set of elements that cover some domain as much as possible).

For polymatroid function (or any monotone non-decreasing function), unconstrained maximization is trivial (take ground set).

Thus, when we do monotone submodular maximization, we either

- Find the maximum under some constraint
- Find the maximum for a non-polymatroid submodular function
- Do both.
Maximization of Submodular Functions

\[
\max_\mathcal{A} f(\mathcal{A}) = \mathcal{V}
\]

- Submodular maximization is quite useful.
- Applications: sensor placement, facility location, document summarization, or any kind of covering problem (choose a small set of elements that cover some domain as much as possible).
- For polymatroid function (or any monotone non-decreasing function), unconstrained maximization is trivial (take ground set).
- Thus, when we do monotone submodular maximization, we either
  - Find the maximum under some constraint
  - Find the maximum for a non-polymatroid submodular function
  - Do both.
- There is also a sort of dual problem that is often considered together with max, and those are minimum cover problems (to be defined).
The Set Cover Problem

Let $E$ be a ground set and let $E_1, E_2, \ldots, E_m$ be a set of subsets.
The Set Cover Problem

- Let $E$ be a ground set and let $E_1, E_2, \ldots, E_m$ be a set of subsets.
- Let $V = \{1, 2, \ldots, m\}$ be the set of integers.
The Set Cover Problem

- Let $E$ be a ground set and let $E_1, E_2, \ldots, E_m$ be a set of subsets.
- Let $V = \{1, 2, \ldots, m\}$ be the set of integers.
- Define $f : 2^V \to \mathbb{Z}_+$ as $f(X) = |\bigcup_{v \in X} E_v|$
The Set Cover Problem

- Let $E$ be a ground set and let $E_1, E_2, \ldots, E_m$ be a set of subsets.
- Let $V = \{1, 2, \ldots, m\}$ be the set of integers.
- Define $f : 2^V \to \mathbb{Z}_+$ as $f(X) = |\bigcup_{v \in X} E_v|$
- Then $f$ is the set cover function. As we say, $f$ is monotone submodular (a polymatroid).
The Set Cover Problem

- Let $E$ be a ground set and let $E_1, E_2, \ldots, E_m$ be a set of subsets.
- Let $V = \{1, 2, \ldots, m\}$ be the set of integers.
- Define $f: 2^V \to \mathbb{Z}_+$ as $f(X) = |\bigcup_{v \in X} E_v|$
- Then $f$ is the set cover function. As we say, $f$ is monotone submodular (a polymatroid).
- The set cover problem asks for the smallest subset $X$ of $V$ such that $f(X) = |E|$ (smallest subset of the subsets of $E$) where $E$ is still covered. I.e.,

\[
\text{minimize } |X| \text{ subject to } f(X) \geq |E| \quad (12.25)
\]
The Set Cover Problem

- Let $E$ be a ground set and let $E_1, E_2, \ldots, E_m$ be a set of subsets.
- Let $V = \{1, 2, \ldots, m\}$ be the set of integers.
- Define $f : 2^V \rightarrow \mathbb{Z}_+$ as $f(X) = |\bigcup_{v \in X} E_v|$
- Then $f$ is the set cover function. As we say, $f$ is monotone submodular (a polymatroid).

The set cover problem asks for the smallest subset $X$ of $V$ such that $f(X) = |E|$ (smallest subset of the subsets of $E$) where $E$ is still covered. I.e.,

$$\text{minimize} |X| \text{ subject to } f(X) \geq |E| \quad (12.25)$$

- We might wish to use a more general modular function $m(X)$ rather than cardinality $|X|$. 
Let $E$ be a ground set and let $E_1, E_2, \ldots, E_m$ be a set of subsets.

Let $V = \{1, 2, \ldots, m\}$ be the set of integers.

Define $f : 2^V \rightarrow \mathbb{Z}_+$ as $f(X) = |\bigcup_{v \in X} E_v|$

Then $f$ is the set cover function. As we say, $f$ is monotone submodular (a polymatroid).

The set cover problem asks for the smallest subset $X$ of $V$ such that $f(X) = |E|$ (smallest subset of the subsets of $E$) where $E$ is still covered. I.e.,

$$\text{minimize } |X| \text{ subject to } f(X) \geq |E| \quad (12.25)$$

We might wish to use a more general modular function $m(X)$ rather than cardinality $|X|$.

This problem is NP-hard, and Feige in 1998 showed that it cannot be approximated with a ratio better than $(1 - \epsilon) \log n$ unless NP is slightly superpolynomial ($n^{O(\log \log n)}$).
What About Non-monotone

- So even simple case of cardinality constrained submodular function maximization is NP-hard.
- This will be true of most submodular max (and related) problems.
- Hence, the only hope is approximation algorithms. Question is, what is the tradeoff between running time and approximation quality, and is it possible to get tight bounds (i.e., an algorithm that achieves an approximation ratio, and a proof that one can’t do better than that unless some extremely unlike event were to be true, such as P=NP).
The Max $k$-Cover Problem

- Let $E$ be a ground set and let $E_1, E_2, \ldots, E_m$ be a set of subsets.
The Max $k$-Cover Problem

- Let $E$ be a ground set and let $E_1, E_2, \ldots, E_m$ be a set of subsets.
- Let $V = \{1, 2, \ldots, m\}$ be the set of integers.
The Max $k$-Cover Problem

- Let $E$ be a ground set and let $E_1, E_2, \ldots, E_m$ be a set of subsets.
- Let $V = \{1, 2, \ldots, m\}$ be the set of integers.
- Define $f : 2^V \rightarrow \mathbb{Z}_+$ as $f(X) = |\bigcup_{v \in V} E_v|$
The Max $k$-Cover Problem

- Let $E$ be a ground set and let $E_1, E_2, \ldots, E_m$ be a set of subsets.
- Let $V = \{1, 2, \ldots, m\}$ be the set of integers.
- Define $f : 2^V \rightarrow \mathbb{Z}_+$ as $f(X) = |\bigcup_{v \in V} E_v|$
- Then $f$ is the set cover function. As we saw, $f$ is monotone submodular (a polymatroid).
The Max $k$-Cover Problem

- Let $E$ be a ground set and let $E_1, E_2, \ldots, E_m$ be a set of subsets.
- Let $V = \{1, 2, \ldots, m\}$ be the set of integers.
- Define $f : 2^V \rightarrow \mathbb{Z}_+ \text{ as } f(X) = |\bigcup_{v \in V} E_v|$
- Then $f$ is the set cover function. As we saw, $f$ is monotone submodular (a polymatroid).
- The max $k$ cover problem asks, given a $k$, what sized $k$ set of sets $X$ can we choose that covers the most? I.e., that maximizes $f(X)$ as in:

$$\max f(X) \text{ subject to } |X| \leq k$$ (12.26)
The Max $k$-Cover Problem

- Let $E$ be a ground set and let $E_1, E_2, \ldots, E_m$ be a set of subsets.
- Let $V = \{1, 2, \ldots, m\}$ be the set of integers.
- Define $f : 2^V \rightarrow \mathbb{Z}_+$ as $f(X) = \left| \bigcup_{v \in V} E_v \right|$.
- Then $f$ is the set cover function. As we saw, $f$ is monotone submodular (a polymatroid).
- The max $k$ cover problem asks, given a $k$, what sized $k$ set of sets $X$ can we choose that covers the most? I.e., that maximizes $f(X)$ as in:

$$\max f(X) \text{ subject to } |X| \leq k$$

(12.26)

- This problem is NP-hard, and Feige in 1998 showed that it cannot be approximated with a ratio better than $(1 - 1/e)$. 
Now we are given an arbitrary polymatroid function $f$. 
Now we are given an arbitrary polymatroid function $f$.

Given $k$, goal is: find $A^* \in \arg \max \{ f(A) : |A| \leq k \}$
Now we are given an arbitrary polymatroid function $f$.

Given $k$, goal is: find $A^* \in \text{argmax} \left\{ f(A) : |A| \leq k \right\}$

w.l.o.g., we can find $A^* \in \text{argmax} \left\{ f(A) : |A| = k \right\}$
Now we are given an arbitrary polymatroid function $f$.

Given $k$, goal is: find $A^* \in \text{argmax} \{ f(A) : |A| \leq k \}$

w.l.o.g., we can find $A^* \in \text{argmax} \{ f(A) : |A| = k \}$

An important result by Nemhauser et. al. (1978) states that for normalized ($f(\emptyset) = 0$) monotone submodular functions (i.e., polymatroids) can be approximately maximized using a simple greedy algorithm.
Now we are given an arbitrary polymatroid function $f$.

Given $k$, goal is: find $A^* \in \text{argmax} \{ f(A) : |A| \leq k \}$

w.l.o.g., we can find $A^* \in \text{argmax} \{ f(A) : |A| = k \}$

An important result by Nemhauser et. al. (1978) states that for normalized ($f(\emptyset) = 0$) monotone submodular functions (i.e., polymatroids) can be approximately maximized using a simple greedy algorithm.

Starting with $S_0 = \emptyset$, we repeat the following greedy step for $i = 0 \ldots (k - 1)$:

$$S_{i+1} = S_i \cup \left\{ \text{argmax} f(S_i \cup \{v\}) \right\}$$  \hspace{1cm} (12.27)
A bit more precisely:

**Algorithm 1: The Greedy Algorithm**

1. Set $S_0 \leftarrow \emptyset$;
2. for $i \leftarrow 0 \ldots |E| - 1$ do
3.  Choose $v_i$ as follows:
   $$v_i \in \arg\max_{v \in V \setminus S_i} f(\{v\} \mid S_i) = \arg\max_{v \in V \setminus S_i} f(S_i \cup \{v\})$$
4.  Set $S_{i+1} \leftarrow S_i \cup \{v_i\}$;
• This algorithm has a guarantee
This algorithm has a guarantee

**Theorem 12.6.1**

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

**Theorem 12.6.1**

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

To find $\tilde{A} \in \arg\max \{f(A) : |A| \leq k\}$, we repeat the greedy step until $k = i + 1$: 
Greedy Algorithm for Card. Constrained Submodular Max

- This algorithm has a guarantee

**Theorem 12.6.1**

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

- To find $A^* \in \arg\max \{ f(A) : |A| \leq k \}$, we repeat the greedy step until $k = i + 1$:
- Again, since this generalizes max $k$-cover, Feige (1998) showed that this can’t be improved. Unless $P = NP$, no polynomial time algorithm can do better than $(1 - 1/e + \epsilon)$ for any $\epsilon > 0$. 
The Greedy Algorithm: $1 - 1/e$ intuition.

- At step $i < k$, greedy chooses $v_i$ to maximize $f(v|S_i)$.

\[ f(v; S_i) \geq f(v|S_i), \quad v \notin V \setminus S_i \]
The Greedy Algorithm: $1 - 1/e$ intuition.

- At step $i < k$, greedy chooses $v_i$ to maximize $f(v|S_i)$.

- Let $S^*$ be optimal solution (of size $k$) and $\text{OPT} = f(S^*)$. 

Equation (12.28) will show that

$$\text{OPT} f(S_i) \leq \frac{1}{1 - 1/e} \cdot \text{OPT} f(S_k)$$

Equation (12.37) will show that

$$\text{OPT} f(S_i + v|S_i) = \frac{1}{k} \cdot \text{OPT} f(S_i)$$
The Greedy Algorithm: $1 - 1/e$ intuition.

- At step $i < k$, greedy chooses $v_i$ to maximize $f(v|S_i)$.
- Let $S^*$ be optimal solution (of size $k$) and $OPT = f(S^*)$. By submodularity, we will show:

\[
\exists v \in V \setminus S_i : f(v|S_i) = f(S_i + v|S_i) \geq \frac{1}{k}(OPT - f(S_i)) \tag{12.28}
\]
The Greedy Algorithm: \(1 - 1/e\) intuition.

- At step \(i < k\), greedy chooses \(v_i\) to maximize \(f(v|S_i)\).
- Let \(S^*\) be optimal solution (of size \(k\)) and \(\text{OPT} = f(S^*)\). By submodularity, we will show:

\[
\exists v \in V \setminus S_i : f(v|S_i) = f(S_i + v|S_i) \geq \frac{1}{k}(\text{OPT} - f(S_i)) \quad (12.28)
\]
The Greedy Algorithm: $1 - 1/e$ intuition.

- At step $i < k$, greedy chooses $v_i$ to maximize $f(v|S_i)$.
- Let $S^*$ be optimal solution (of size $k$) and $\text{OPT} = f(S^*)$. By submodularity, we will show:

$$\exists v \in V \setminus S_i : f(v|S_i) = f(S_i + v|S_i) \geq \frac{1}{k} (\text{OPT} - f(S_i))$$

(12.28)
The Greedy Algorithm: $1 - 1/e$ intuition.

- At step $i < k$, greedy chooses $v_i$ to maximize $f(v|S_i)$.
- Let $S^*$ be optimal solution (of size $k$) and $OPT = f(S^*)$. By submodularity, we will show:

$$\exists v \in V \setminus S_i : f(v|S_i) = f(S_i + v|S_i) \geq \frac{1}{k}(OPT - f(S_i))$$  \hspace{1cm} (12.28)

Equation (12.37) will show that
Equation (12.28) $\Rightarrow$:

$$OPT - f(S_{i+1}) \leq (1 - 1/k)(OPT - f(S_i))$$
$$\Rightarrow OPT - f(S_k) \leq (1 - 1/k)^k OPT$$
$$\leq 1/eOPT$$
$$\Rightarrow OPT(1 - 1/e) \leq f(S_k)$$
Given non-negative monotone submodular function \( f : 2^V \rightarrow \mathbb{R}_+ \), define \( \{S_i\}_{i \geq 0} \) to be the chain formed by the greedy algorithm (Eqn. (12.27)). Then for all \( k, \ell \in \mathbb{Z}_{++} \), we have:

\[
f(S_\ell) \geq (1 - e^{-\ell/k}) \max_{S : |S| \leq k} f(S) \tag{12.29}
\]

and in particular, for \( \ell = k \), we have \( f(S_k) \geq (1 - 1/e) \max_{S : |S| \leq k} f(S) \).
Theorem 12.6.2 (Nemhauser et al. 1978)

Given non-negative monotone submodular function \( f : 2^V \to \mathbb{R}_+ \), define \( \{S_i\}_{i \geq 0} \) to be the chain formed by the greedy algorithm (Eqn. (12.27)). Then for all \( k, \ell \in \mathbb{Z}_{++} \), we have:

\[
f(S_\ell) \geq (1 - e^{-\ell/k}) \max_{S:|S| \leq k} f(S)
\]

(12.29)

and in particular, for \( \ell = k \), we have \( f(S_k) \geq (1 - 1/e) \max_{S:|S| \leq k} f(S) \).

- \( k \): is size of optimal set, i.e., \( \text{OPT} = f(S^*) \) with \( |S^*| = k \)
Given non-negative monotone submodular function $f : 2^V \to \mathbb{R}_+$, define \( \{S_i\}_{i \geq 0} \) to be the chain formed by the greedy algorithm (Eqn. (12.27)). Then for all $k, \ell \in \mathbb{Z}_{++}$, we have:

$$f(S_\ell) \geq (1 - e^{-\ell/k}) \max_{S : |S| \leq k} f(S)$$  \hspace{1cm} (12.29)

and in particular, for $\ell = k$, we have $f(S_k) \geq (1 - 1/e) \max_{S : |S| \leq k} f(S)$.

- $k$ is size of optimal set, i.e., $\text{OPT} = f(S^*)$ with $|S^*| = k$.
- $\ell$ is size of set we are choosing (i.e., we choose $S_\ell$ from greedy chain).
Cardinality Constrained Polymatroid Max Theorem

Theorem 12.6.2 (Nemhauser et al. 1978)

Given non-negative monotone submodular function \( f : 2^V \to \mathbb{R}_+ \), define \( \{S_i\}_{i \geq 0} \) to be the chain formed by the greedy algorithm (Eqn. (12.27)). Then for all \( k, \ell \in \mathbb{Z}_{++} \), we have:

\[
f(S_\ell) \geq (1 - e^{-\ell/k}) \max_{S: |S| \leq k} f(S)
\]  

(12.29)

and in particular, for \( \ell = k \), we have \( f(S_k) \geq (1 - 1/e) \max_{S: |S| \leq k} f(S) \).

- \( k \) is size of optimal set, i.e., \( \text{OPT} = f(S^*) \) with \( |S^*| = k \)
- \( \ell \) is size of set we are choosing (i.e., we choose \( S_\ell \) from greedy chain).
- Bound is how well does \( S_\ell \) (of size \( \ell \)) do relative to \( S^* \), the optimal set of size \( k \).
Theorem 12.6.2 (Nemhauser et al. 1978)

Given non-negative monotone submodular function \( f : 2^V \rightarrow \mathbb{R}_+ \), define \( \{S_i\}_{i \geq 0} \) to be the chain formed by the greedy algorithm (Eqn. (12.27)). Then for all \( k, \ell \in \mathbb{Z}_+ \), we have:

\[
f(S_\ell) \geq (1 - e^{-\ell/k}) \max_{S:|S| \leq k} f(S)
\]

(12.29)

and in particular, for \( \ell = k \), we have \( f(S_k) \geq (1 - 1/e) \max_{S:|S| \leq k} f(S) \).

- \( k \) is size of optimal set, i.e., \( \text{OPT} = f(S^*) \) with \( |S^*| = k \).
- \( \ell \) is size of set we are choosing (i.e., we choose \( S_\ell \) from greedy chain).
- Bound is how well does \( S_\ell \) (of size \( \ell \)) do relative to \( S^* \), the optimal set of size \( k \).
- Intuitively, bound should get worse when \( \ell < k \) and get better when \( \ell > k \).
Cardinality Constrained Polymatroid Max Theorem

Proof of Theorem 12.6.2.

Fix $(\text{number of items greedy will choose})$ and $k$ (size of optimal set to compare against). Set $S^\ast \in \arg\max \{ f(S) : |S| \leq k \}$. w.l.o.g. assume $|S^\ast| = k$. Order $S^\ast = (v^\ast_1, v^\ast_2, ..., v^\ast_k)$ arbitrarily. Let $S_i = (v_1, v_2, ..., v_i)$ be the greedy order chain chosen by the algorithm, for $i \in \{1, 2, ..., k\}$. Then the following inequalities follow:
Proof of Theorem 12.6.2.

- Fix \( \ell \) (number of items greedy will chose) and \( k \) (size of optimal set to compare against).
Proof of Theorem 12.6.2.

- Fix \( \ell \) (number of items greedy will choose) and \( k \) (size of optimal set to compare against).
- Set \( S^* \in \arg\max \{ f(S) : |S| \leq k \} \)
Proof of Theorem 12.6.2.

- Fix $\ell$ (number of items greedy will chose) and $k$ (size of optimal set to compare against).
- Set $S^* \in \text{argmax} \{ f(S) : |S| \leq k \}$
- w.l.o.g. assume $|S^*| = k$. 

...
Proof of Theorem 12.6.2.

- Fix \( \ell \) (number of items greedy will chose) and \( k \) (size of optimal set to compare against).
- Set \( S^* \in \text{argmax} \{ f(S) : |S| \leq k \} \)
- w.l.o.g. assume \( |S^*| = k \).
- Order \( S^* = (v_1^*, v_2^*, \ldots, v_k^*) \) arbitrarily.
Proof of Theorem 12.6.2.

- Fix \( \ell \) (number of items greedy will choose) and \( k \) (size of optimal set to compare against).
- Set \( S^* \in \text{argmax} \{ f(S) : |S| \leq k \} \)
- w.l.o.g. assume \( |S^*| = k \).
- Order \( S^* = (v_1^*, v_2^*, \ldots, v_k^*) \) arbitrarily.
- Let \( S_i = (v_1, v_2, \ldots, v_i) \) be the greedy order chain chosen by the algorithm, for \( i \in \{1, 2, \ldots, \ell\} \).
Proof of Theorem 12.6.2.

- Fix $\ell$ (number of items greedy will chose) and $k$ (size of optimal set to compare against).
- Set $S^* \in \arg\max \{f(S) : |S| \leq k\}$
- w.l.o.g. assume $|S^*| = k$.
- Order $S^* = (v_1^*, v_2^*, \ldots, v_k^*)$ arbitrarily.
- Let $S_i = (v_1, v_2, \ldots, v_i)$ be the greedy order chain chosen by the algorithm, for $i \in \{1, 2, \ldots, \ell\}$.
- Then the following inequalities (on the next slide) follow:
... proof of Theorem 12.6.2 cont.
Cardinality Constrained Polymatroid Max Theorem

... proof of Theorem 12.6.2 cont.

- For all $i < \ell$, we have

$$f(S^*)$$
For all $i < \ell$, we have

$$f(S^*) \leq f(S^* \cup S_i)$$
... proof of Theorem 12.6.2 cont.

For all $i < \ell$, we have

$$f(S^*) \leq f(S^* \cup S_i) = f(S_i) + f(S^* | S_i)$$  \hfill (12.30)
... proof of Theorem 12.6.2 cont.

For all $i < \ell$, we have

$$f(S^*) \leq f(S^* \cup S_i) = f(S_i) + f(S^* | S_i)$$  \hspace{1cm} (12.30)

$$= f(S_i) + \sum_{j=1}^{K} f(v_j^* | S_i \cup \{v_1^*, v_2^*, \ldots, v_{j-1}^*\})$$  \hspace{1cm} (12.31)
Cardinality Constrained Polymatroid Max Theorem

... proof of Theorem 12.6.2 cont.

For all $i < \ell$, we have

$$f(S^*) \leq f(S^* \cup S_i) = f(S_i) + f(S^* | S_i)$$

(12.30)

$$= f(S_i) + \sum_{j=1}^{k} f(v_j^* | S_i \cup \{v_1^*, v_2^*, \ldots, v_{j-1}^*\})$$

(12.31)

$$\leq f(S_i) + \sum_{v \in S^*} f(v | S_i)$$

(12.32)
For all $i < \ell$, we have

\[
\begin{align*}
f(S^*) & \leq f(S^* \cup S_i) = f(S_i) + f(S^* \mid S_i) \\
&= f(S_i) + \sum_{j=1}^{k} f(v_j^* \mid S_i \cup \{v_1^*, v_2^*, \ldots, v_{j-1}^*\}) \\
&\leq f(S_i) + \sum_{v \in S^*} f(v \mid S_i) \\
&\leq f(S_i) + \sum_{v \in S^*} f(v_{i+1} \mid S_i)
\end{align*}
\]
For all $i < \ell$, we have

$$f(S^*) \leq f(S^* \cup S_i) = f(S_i) + f(S^*|S_i)$$

(12.30)

$$= f(S_i) + \sum_{j=1}^{k} f(v_j^*|S_i \cup \{v_1^*, v_2^*, \ldots, v_{j-1}^*\})$$

(12.31)

$$\leq f(S_i) + \sum_{v \in S^*} f(v|S_i)$$

(12.32)

$$\leq f(S_i) + \sum_{v \in S^*} f(v_{i+1}|S_i) = f(S_i) + \sum_{v \in S^*} f(S_{i+1}|S_i)$$

(12.33)

\[\ldots\]
... proof of Theorem 12.6.2 cont.

For all \( i < \ell \), we have

\[
\begin{align*}
f(S^*) &\leq f(S^* \cup S_i) = f(S_i) + f(S^* \mid S_i) \\
&= f(S_i) + \sum_{j=1}^{k} f(v_j^* \mid S_i \cup \{v_1^*, v_2^*, \ldots, v_{j-1}^*\}) \\
&\leq f(S_i) + \sum_{v \in S^*} f(v \mid S_i) \\
&\leq f(S_i) + \sum_{v \in S^*} f(v_{i+1} \mid S_i) = f(S_i) + \sum_{v \in S^*} f(S_{i+1} \mid S_i) \\
&= f(S_i) + kf(S_{i+1} \mid S_i)
\end{align*}
\]

(12.30)
Proof of Theorem 12.6.2 cont.

For all $i < \ell$, we have

\[
    f(S^*) \leq f(S^* \cup S_i) = f(S_i) + f(S^* | S_i)
\]

\[
    = f(S_i) + \sum_{j=1}^{k} f(v_j^* | S_i \cup \{v_1^*, v_2^*, \ldots, v_{j-1}^*\})
\]

\[
    \leq f(S_i) + \sum_{v \in S^*} f(v|S_i)
\]

\[
    \leq f(S_i) + \sum_{v \in S^*} f(v_{i+1}|S_i) = f(S_i) + \sum_{v \in S^*} f(S_{i+1}|S_i)
\]

\[
    = f(S_i) + kf(S_{i+1}|S_i)
\]

Therefore, we have Equation 12.28, i.e.,:

\[
    f(S^*) - f(S_i) \leq kf(S_{i+1}|S_i) = k(f(S_{i+1}) - f(S_i))
\]
Define $i, f(S_i) \Rightarrow f(S_i+1)$, giving

$$i \leq k (i+1) \quad (12.36)$$

or

$$i+1 \leq (1 - 1/k) i \quad (12.37)$$

The relationship between $0$ and $\`$ is then

$$l \leq (1 - 1/k) \quad (12.38)$$

Now, $0 = f(S^\ast) \Rightarrow f(S^\ast)$ since $f_0$.

Also, by variational bound $1^T x \leq e^T x$ for $x \in \mathbb{R}$, we have

$$\` \leq (1 - 1/k) f(S^\ast) \quad (12.39)$$
Define $\delta_i \triangleq f(S^*) - f(S_i)$, so $\delta_i - \delta_{i+1} = f(S_{i+1}) - f(S_i)$.
Define $\delta_i \triangleq f(S^*) - f(S_i)$, so $\delta_i - \delta_{i+1} = f(S_{i+1}) - f(S_i)$, giving
\[
\delta_i \leq k(\delta_i - \delta_{i+1}) \tag{12.36}
\]
or
Define $\delta_i \triangleq f(S^*) - f(S_i)$, so $\delta_i - \delta_{i+1} = f(S_{i+1}) - f(S_i)$, giving
\[ \delta_i \leq k(\delta_i - \delta_{i+1}) \tag{12.36} \]
or
\[ \delta_{i+1} \leq (1 - \frac{1}{k})\delta_i \tag{12.37} \]
Cardinality Constrained Polymatroid Max Theorem

...proof of Theorem 12.6.2 cont.

- Define \( \delta_i \triangleq f(S^*) - f(S_i) \), so \( \delta_i - \delta_{i+1} = f(S_{i+1}) - f(S_i) \), giving
  \[
  \delta_i \leq k(\delta_i - \delta_{i+1})
  \]  
  (12.36)
  or
  \[
  \delta_{i+1} \leq (1 - \frac{1}{k})\delta_i
  \]  
  (12.37)
- The relationship between \( \delta_0 \) and \( \delta_\ell \) is then
  \[
  \delta_\ell \leq (1 - \frac{1}{k})^\ell \delta_0
  \]  
  (12.38)
Define $\delta_i \triangleq f(S^*) - f(S_i)$, so $\delta_i - \delta_{i+1} = f(S_{i+1}) - f(S_i)$, giving
$$\delta_i \leq k(\delta_i - \delta_{i+1})$$
(12.36)

or
$$\delta_{i+1} \leq (1 - \frac{1}{k})\delta_i$$
(12.37)

The relationship between $\delta_0$ and $\delta_\ell$ is then
$$\delta_\ell \leq (1 - \frac{1}{k})^\ell \delta_0$$
(12.38)

Now, $\delta_0 = f(S^*) - f(\emptyset) \leq f(S^*)$ since $f \geq 0$. 
... proof of Theorem 12.6.2 cont.

- Define $\delta_i \triangleq f(S^*) - f(S_i)$, so $\delta_i - \delta_{i+1} = f(S_{i+1}) - f(S_i)$, giving
  
  $\delta_i \leq k(\delta_i - \delta_{i+1})$ \hspace{1cm} (12.36)

  or

  $\delta_{i+1} \leq (1 - \frac{1}{k})\delta_i$ \hspace{1cm} (12.37)

- The relationship between $\delta_0$ and $\delta_{\ell}$ is then

  $\delta_{\ell} \leq (1 - \frac{1}{k})^\ell \delta_0$ \hspace{1cm} (12.38)

- Now, $\delta_0 = f(S^*) - f(\emptyset) \leq f(S^*)$ since $f \geq 0$.

- Also, by variational bound $1 - x \leq e^{-x}$ for $x \in \mathbb{R}$, we have

  $\delta_{\ell} \leq (1 - \frac{1}{k})^\ell \delta_0 \leq e^{-\ell/k} f(S^*)$ \hspace{1cm} (12.39)
When we identify $l = f(S^\ast)$, a bit of rearranging then gives:

$$f(S^\ast) = (1 - e^{-k}) f(S) \tag{12.40}$$

With $k = k$, picking $k$ items, greedy gets $\frac{1}{e} \approx 0.6321$ bound. This means that if $S_k$ is greedy solution of size $k$, and $S^\ast$ is an optimal solution of size $k$, $f(S_k)$ is $\approx 0.6321 f(S^\ast)$.

What if we want to guarantee a solution no worse than $0.95 f(S^\ast)$ where $|S^\ast| = k$? Set $0.95 = (1 - e^{-d/k \ln(1 - 0.999)})$, which gives $d/k \ln(1 - 0.999) = 7$. So solution, in the worst case, quickly gets very good. Typical/practical case is much better.
When we identify $\delta_\ell = f(S^*) - f(S_\ell)$, a bit of rearranging then gives:

$$f(S_\ell) \geq (1 - e^{-\ell/k}) f(S^*)$$

(12.40)
...proof of Theorem 12.6.2 cont.

When we identify $\delta_l = f(S^*) - f(S_\ell)$, a bit of rearranging then gives:

$$f(S_\ell) \geq (1 - e^{-\ell/k}) f(S^*)$$

(12.40)

With $\ell = k$, when picking $k$ items, greedy gets $(1 - 1/e) \approx 0.6321$ bound. This means that if $S_k$ is greedy solution of size $k$, and $S^*$ is an optimal solution of size $k$, $f(S_k) \geq (1 - 1/e) f(S^*) \approx 0.6321 f(S^*)$. 
... proof of Theorem 12.6.2 cont.

- When we identify \( \delta_{\ell} = f(S^\ast) - f(S_{\ell}) \), a bit of rearranging then gives:

\[
f(S_{\ell}) \geq (1 - e^{-\ell/k}) f(S^\ast)
\]  

(12.40)

- With \( \ell = k \), when picking \( k \) items, greedy gets \( (1 - 1/e) \approx 0.6321 \) bound. This means that if \( S_k \) is greedy solution of size \( k \), and \( S^\ast \) is an optimal solution of size \( k \), \( f(S_k) \geq (1 - 1/e) f(S^\ast) \approx 0.6321 f(S^\ast) \).

- What if we want to guarantee a solution no worse than \( .95 f(S^\ast) \) where \( |S^\ast| = k \)?
Cardinality Constrained Polymatroid Max Theorem

... proof of Theorem 12.6.2 cont.

- When we identify $\delta_l = f(S^*) - f(S_l)$, a bit of rearranging then gives:

$$f(S_l) \geq (1 - e^{-\ell/k})f(S^*) \quad (12.40)$$

- With $\ell = k$, when picking $k$ items, greedy gets $(1 - 1/e) \approx 0.6321$ bound. This means that if $S_k$ is greedy solution of size $k$, and $S^*$ is an optimal solution of size $k$, $f(S_k) \geq (1 - 1/e)f(S^*) \approx 0.6321f(S^*)$.

- What if we want to guarantee a solution no worse than $0.95f(S^*)$ where $|S^*| = k$? Set $0.95 = (1 - e^{-\ell/k})$, which gives $\ell = \lceil -k \ln(1 - 0.95) \rceil = 4k$. 

When we identify $\delta_l = f(S^*) - f(S_{\ell})$, a bit of rearranging then gives:

$$f(S_{\ell}) \geq (1 - e^{-\ell/k}) f(S^*)$$  \hfill (12.40)

With $\ell = k$, when picking $k$ items, greedy gets $(1 - 1/e) \approx 0.6321$ bound. This means that if $S_k$ is greedy solution of size $k$, and $S^*$ is an optimal solution of size $k$, $f(S_k) \geq (1 - 1/e) f(S^*) \approx 0.6321 f(S^*)$.

What if we want to guarantee a solution no worse than $.95 f(S^*)$ where $|S^*| = k$? Set $0.95 = (1 - e^{-\ell/k})$, which gives $\ell = \lceil -k \ln(1 - 0.95) \rceil = 4k$. And $\lceil -\ln(1 - 0.999) \rceil = 7$. 
Cardinality Constrained Polymatroid Max Theorem

...proof of Theorem 12.6.2 cont.

- When we identify $\delta_l = f(S^*) - f(S_\ell)$, a bit of rearranging then gives:

  $$f(S_\ell) \geq (1 - e^{-\ell/k}) f(S^*) \quad (12.40)$$

- With $\ell = k$, when picking $k$ items, greedy gets $(1 - 1/e) \approx 0.6321$ bound. This means that if $S_k$ is greedy solution of size $k$, and $S^*$ is an optimal solution of size $k$, $f(S_k) \geq (1 - 1/e) f(S^*) \approx 0.6321 f(S^*)$.

- What if we want to guarantee a solution no worse than $0.95 f(S^*)$ where $|S^*| = k$? Set $0.95 = (1 - e^{-\ell/k})$, which gives

  $$\ell = \lceil -k \ln(1 - 0.95) \rceil = 4k.$$  And $\lceil -\ln(1 - 0.999) \rceil = 7$.

- So solution, in the worst case, quickly gets very good. Typical/practical case is much better.
Greedy running time

- Greedy computes a new maximum $n = |V|$ times, and each maximum computation requires $O(n)$ comparisons, leading to $O(n^2)$ computation for greedy.
Greedy running time

- Greedy computes a new maximum \( n = |V| \) times, and each maximum computation requires \( O(n) \) comparisons, leading to \( O(n^2) \) computation for greedy.

- This is the best we can do for arbitrary functions, but \( O(n^2) \) is not practical to some.
Greedy running time

- Greedy computes a new maximum $n = |V|$ times, and each maximum computation requires $O(n)$ comparisons, leading to $O(n^2)$ computation for greedy.
- This is the best we can do for arbitrary functions, but $O(n^2)$ is not practical to some.
- Greedy can be made much faster in practice by a simple strategy made possible, once again, via the use of submodularity.
Greedy running time

- Greedy computes a new maximum $n = |V|$ times, and each maximum computation requires $O(n)$ comparisons, leading to $O(n^2)$ computation for greedy.

- This is the best we can do for arbitrary functions, but $O(n^2)$ is not practical to some.

- Greedy can be made much faster in practice by a simple strategy made possible, once again, via the use of submodularity.

- This is called Minoux’s 1977 Accelerated Greedy strategy (and has been rediscovered a few times, e.g., “Lazy greedy”), and runs much faster while still producing same answer.
Greedy running time

- Greedy computes a new maximum $n = |V|$ times, and each maximum computation requires $O(n)$ comparisons, leading to $O(n^2)$ computation for greedy.
- This is the best we can do for arbitrary functions, but $O(n^2)$ is not practical to some.
- Greedy can be made much faster in practice by a simple strategy made possible, once again, via the use of submodularity.
- This is called Minoux’s 1977 Accelerated Greedy strategy (and has been rediscovered a few times, e.g., “Lazy greedy”), and runs much faster while still producing same answer.
- We describe it next:
At stage $i$ in the algorithm, we have a set of gains $f(v|S_i)$ for all $v \notin S_i$. Store these values $\alpha_v \leftarrow f(v|S_i)$ in sorted priority queue.
Minoux’s Accelerated Greedy for Submodular Functions

- At stage $i$ in the algorithm, we have a set of gains $f(v|S_i)$ for all $v \notin S_i$. Store these values $\alpha_v \leftarrow f(v|S_i)$ in sorted priority queue.
- Priority queue, $O(1)$ to find max, $O(\log n)$ to insert in right place.
Minoux’s Accelerated Greedy for Submodular Functions

- At stage $i$ in the algorithm, we have a set of gains $f(v|S_i)$ for all $v \notin S_i$. Store these values $\alpha_v \leftarrow f(v|S_i)$ in sorted priority queue.
- Priority queue, $O(1)$ to find max, $O(\log n)$ to insert in right place.
- Once we choose a max $v$, then set $S_{i+1} \leftarrow S_i + v$. 

Minoux’s Accelerated Greedy for Submodular Functions

- At stage $i$ in the algorithm, we have a set of gains $f(v|S_i)$ for all $v \notin S_i$. Store these values $\alpha_v \leftarrow f(v|S_i)$ in sorted priority queue.
- Priority queue, $O(1)$ to find max, $O(\log n)$ to insert in right place.
- Once we choose a max $v$, then set $S_{i+1} \leftarrow S_i + v$.
- For $v \notin S_{i+1}$ we have $f(v|S_{i+1}) \leq f(v|S_i)$ by submodularity.
Minoux’s Accelerated Greedy for Submodular Functions

- At stage $i$ in the algorithm, we have a set of gains $f(v|S_i)$ for all $v \notin S_i$. Store these values $\alpha_v \leftarrow f(v|S_i)$ in sorted priority queue.
- Priority queue, $O(1)$ to find max, $O(\log n)$ to insert in right place.
- Once we choose a max $v$, then set $S_{i+1} \leftarrow S_i + v$.
- For $v \notin S_{i+1}$ we have $f(v|S_{i+1}) \leq f(v|S_i)$ by submodularity.
- Therefore, if we find a $v'$ such that $f(v'|S_{i+1}) \geq \alpha_v$ for all $v \neq v'$, then since

$$f(v'|S_{i+1}) \geq \alpha_v = f(v|S_i) \geq f(v|S_{i+1}) \quad (12.41)$$

we have the true max, and we need not re-evaluate gains of other elements again.
At stage $i$ in the algorithm, we have a set of gains $f(v|S_i)$ for all $v \notin S_i$. Store these values $\alpha_v \leftarrow f(v|S_i)$ in sorted priority queue.

Priority queue, $O(1)$ to find max, $O(\log n)$ to insert in right place.

Once we choose a max $v$, then set $S_{i+1} \leftarrow S_i + v$.

For $v \notin S_{i+1}$ we have $f(v|S_{i+1}) \leq f(v|S_i)$ by submodularity.

Therefore, if we find a $v'$ such that $f(v'|S_{i+1}) \geq \alpha_v$ for all $v \neq v'$, then since

$$f(v'|S_{i+1}) \geq \alpha_v \geq f(v|S_i) \geq f(v|S_{i+1})$$

we have the true max, and we need not re-evaluate gains of other elements again.

Strategy is: find the $\arg\max_{v' \in V \setminus S_{i+1}} \alpha_{v'}$, and then compute the real $f(v'|S_{i+1})$. If it is greater than all other $\alpha_{v'}$‘s then that’s the next greedy step. Otherwise, replace $\alpha_{v'}$ with its real value, resort, and repeat.
Minoux’s Accelerated Greedy for Submodular Functions

- Minoux’s algorithm is exact, in that it has the same guarantees as does the $O(n^2)$ greedy Algorithm 2 (this means it will return either the same answers, or answers that have the $1 - 1/e$ guarantee).
Minoux’s Accelerated Greedy for Submodular Functions

- Minoux’s algorithm is exact, in that it has the same guarantees as does the $O(n^2)$ greedy Algorithm 2 (this means it will return either the same answers, or answers that have the $1 - 1/e$ guarantee).

- In practice: Minoux’s trick has enormous speedups ($\approx 700\times$) over the standard greedy procedure due to reduced function evaluations and use of good data structures (priority queue).
Minoux’s Accelerated Greedy for Submodular Functions

- Minoux’s algorithm is exact, in that it has the same guarantees as does the $O(n^2)$ greedy Algorithm 2 (this means it will return either the same answers, or answers that have the $1 - 1/e$ guarantee).
- In practice: Minoux’s trick has enormous speedups ($\approx 700\times$) over the standard greedy procedure due to reduced function evaluations and use of good data structures (priority queue).
- When choosing a of size $k$, naïve greedy algorithm is $O(nk)$ but accelerated variant at the very best does $O(n + k)$, so this limits the speedup.
Minoux’s algorithm is exact, in that it has the same guarantees as does the \( O(n^2) \) greedy Algorithm 2 (this means it will return either the same answers, or answers that have the \( 1 - 1/e \) guarantee).

In practice: Minoux’s trick has enormous speedups (\( \approx 700 \times \) ) over the standard greedy procedure due to reduced function evaluations and use of good data structures (priority queue).

When choosing a of size \( k \), naïve greedy algorithm is \( O(nk) \) but accelerated variant at the very best does \( O(n + k) \), so this limits the speedup.

Algorithm has been rediscovered (I think) independently (CELF - cost-effective lazy forward selection, Leskovec et al., 2007)
Minoux’s Accelerated Greedy for Submodular Functions

- Minoux’s algorithm is exact, in that it has the same guarantees as does the $O(n^2)$ greedy Algorithm 2 (this means it will return either the same answers, or answers that have the $1 - 1/e$ guarantee).
- In practice: Minoux’s trick has enormous speedups ($\approx 700 \times$) over the standard greedy procedure due to reduced function evaluations and use of good data structures (priority queue).
- When choosing a of size $k$, naïve greedy algorithm is $O(nk)$ but accelerated variant at the very best does $O(n + k)$, so this limits the speedup.
- Algorithm has been rediscovered (I think) independently (CELF - cost-effective lazy forward selection, Leskovec et al., 2007)
- Can be used used for “big data” sets (e.g., social networks, selecting blogs of greatest influence, document summarization, etc.).
Priority Queue

- Use a priority queue $Q$ as a data structure: operations include:

1. Insert an item $(v, \rho)$ into queue, with $v \in V$ and $\rho \in \mathbb{R}$.

   \begin{equation}
   \text{insert}(Q, (v, \rho))
   \end{equation}

2. Pop the item $(v, \rho)$ with maximum value $\rho$ from the queue.

   \begin{equation}
   (v, \rho) \text{pop}(Q)
   \end{equation}

3. Query the value of the max item in the queue.

   \begin{equation}
   \max(Q)
   \end{equation}

On next slide, we call a popped item “fresh” if the value $(v, \rho)$ popped has the correct value $\rho = f(v | S_i)$. Use extra “bit” to store this info. If a popped item is fresh, it must be the maximum — this can happen if, at given iteration, $v$ was first popped and neither fresh nor maximum so placed back in the queue, and it then percolates back to the top at which point it is fresh — thereby avoid extra queue check.
Use a priority queue $Q$ as a data structure: operations include:

- Insert an item $(v, \alpha)$ into queue, with $v \in V$ and $\alpha \in \mathbb{R}$.

\begin{equation}
\text{INSERT}(Q, (v, \alpha))
\end{equation}
Priority Queue

Use a priority queue $Q$ as a data structure: operations include:

- Insert an item $(v, \alpha)$ into queue, with $v \in V$ and $\alpha \in \mathbb{R}$.
  
  $$\text{INSERT}(Q, (v, \alpha))$$  \hfill (12.42)

- Pop the item $(v, \alpha)$ with maximum value $\alpha$ off the queue.
  
  $$(v, \alpha) \leftarrow \text{POP}(Q)$$  \hfill (12.43)
Priority Queue

- Use a priority queue $Q$ as a data structure: operations include:
  - Insert an item $(v, \alpha)$ into queue, with $v \in V$ and $\alpha \in \mathbb{R}$.
    
    $\text{INSERT}(Q, (v, \alpha))$  
    \hspace{1cm} (12.42)

  - Pop the item $(v, \alpha)$ with maximum value $\alpha$ off the queue.
    
    $(v, \alpha) \leftarrow \text{POP}(Q)$  
    \hspace{1cm} (12.43)

  - Query the value of the max item in the queue
    
    $\text{MAX}(Q) \in \mathbb{R}$  
    \hspace{1cm} (12.44)
Priority Queue

- Use a priority queue $Q$ as a data structure: operations include:
  - Insert an item $(v, \alpha)$ into queue, with $v \in V$ and $\alpha \in \mathbb{R}$.
    \[ \text{INSERT}(Q, (v, \alpha)) \] (12.42)
  - Pop the item $(v, \alpha)$ with maximum value $\alpha$ off the queue.
    \[ (v, \alpha) \leftarrow \text{POP}(Q) \] (12.43)
  - Query the value of the max item in the queue
    \[ \text{MAX}(Q) \in \mathbb{R} \] (12.44)

- On next slide, we call a popped item “fresh” if the value $(v, \alpha)$ popped has the correct value $\alpha = f(v|S_i)$. Use extra “bit” to store this info.
Priority Queue

- Use a priority queue $Q$ as a data structure: operations include:
  - Insert an item $(v, \alpha)$ into queue, with $v \in V$ and $\alpha \in \mathbb{R}$.
    \[
    \text{INSERT}(Q, (v, \alpha))
    \] (12.42)
  - Pop the item $(v, \alpha)$ with maximum value $\alpha$ off the queue.
    \[
    (v, \alpha) \leftarrow \text{POP}(Q)
    \] (12.43)
  - Query the value of the max item in the queue
    \[
    \text{MAX}(Q) \in \mathbb{R}
    \] (12.44)

- On next slide, we call a popped item “fresh” if the value $(v, \alpha)$ popped has the correct value $\alpha = f(v|S_i)$. Use extra “bit” to store this info.

- If a popped item is fresh, it must be the maximum — this can happen if, at given iteration, $v$ was first popped and neither fresh nor maximum so placed back in the queue, and it then percolates back to the top at which point it is fresh — thereby avoid extra queue check.
Algorithm 2: Minoux’s Accelerated Greedy Algorithm

1. Set $S_0 \leftarrow \emptyset$; $i \leftarrow 0$; Initialize priority queue $Q$;
2. for $v \in E$ do
   3. INSERT($Q$, $f(v)$)
3. repeat
   4. $(v, \alpha) \leftarrow \text{POP}(Q)$;
   5. if $\alpha$ not “fresh” then
      6. recomputes $\alpha \leftarrow f(v|S_i)$
   7. if (popped $\alpha$ in line 5 was “fresh”) OR ($\alpha \geq \text{MAX}(Q)$) then
      8. Set $S_{i+1} \leftarrow S_i \cup \{v\}$;
      9. $i \leftarrow i + 1$
   10. else
      11. INSERT($Q$, $(v, \alpha)$)
4. until $i = |E|$;
Minimum Submodular Cover

- Given polymatroid $f$, goal is to find a covering set of minimum cost:

  $$S^* \in \underset{S \subseteq V}{\operatorname{argmin}} |S| \text{ such that } f(S) \geq \alpha$$  \hspace{1cm} (12.45)

  where $\alpha$ is a “cover” requirement.
Minimum Submodular Cover

- Given polymatroid $f$, goal is to find a covering set of minimum cost:

$$S^* \in \arg\min_{S \subseteq V} |S| \text{ such that } f(S) \geq \alpha$$

(12.45)

where $\alpha$ is a “cover” requirement.

- Normally take $\alpha = f(V)$ but defining $f'(A) = \min \{ f(A), \alpha \}$ we can take any $\alpha$. Hence, we have equivalent formulation:

$$S^* \in \arg\min_{S \subseteq V} |S| \text{ such that } f'(S) \geq f'(V)$$

(12.46)
Minimum Submodular Cover

- Given polymatroid \( f \), goal is to find a covering set of minimum cost:
  \[
  S^* \in \arg\min_{S \subseteq V} |S| \text{ such that } f(S) \geq \alpha
  \]  \hspace{1cm} (12.45)

  where \( \alpha \) is a “cover” requirement.

- Normally take \( \alpha = f(V) \) but defining \( f'(A) = \min \{ f(A), \alpha \} \) we can take any \( \alpha \). Hence, we have equivalent formulation:
  \[
  S^* \in \arg\min_{S \subseteq V} |S| \text{ such that } f'(S) \geq f'(V)
  \]  \hspace{1cm} (12.46)

- Note that this immediately generalizes standard set cover, in which case \( f(A) \) is the cardinality of the union of sets indexed by \( A \).
Minimum Submodular Cover

- Given polymatroid $f$, goal is to find a covering set of minimum cost:
  \[ S^* \in \arg\min_S |S| \text{ such that } f(S) \geq \alpha \]  
  (12.45)

  where $\alpha$ is a “cover” requirement.

- Normally take $\alpha = f(V)$ but defining $f'(A) = \min \{ f(A), \alpha \}$ we can take any $\alpha$. Hence, we have equivalent formulation:
  \[ S^* \in \arg\min_S |S| \text{ such that } f'(S) \geq f'(V) \]  
  (12.46)

- Note that this immediately generalizes standard set cover, in which case $f(A)$ is the cardinality of the union of sets indexed by $A$.

- Algorithm: Pick the first $S_i$ chosen by aforementioned greedy algorithm such that $f(S_i) \geq \alpha$. 

Minimum Submodular Cover

- Given polymatroid $f$, goal is to find a covering set of minimum cost:

$$S^* \in \arg\min \left\{ |S| \mid S \subseteq V \text{ such that } f(S) \geq \alpha \right\} \tag{12.45}$$

where $\alpha$ is a “cover” requirement.

- Normally take $\alpha = f(V)$ but defining $f'(A) = \min \{ f(A), \alpha \}$ we can take any $\alpha$. Hence, we have equivalent formulation:

$$S^* \in \arg\min \left\{ |S| \mid S \subseteq V \text{ such that } f'(S) \geq f'(V) \right\} \tag{12.46}$$

- Note that this immediately generalizes standard set cover, in which case $f(A)$ is the cardinality of the union of sets indexed by $A$.

- Algorithm: Pick the first $S_i$ chosen by aforementioned greedy algorithm such that $f(S_i) \geq \alpha$.

- For integer valued $f$, this greedy algorithm an $O(\log(\max_{s \in V} f(\{s\})))$ approximation. Set cover is hard to approximate with a factor better than $(1 - \epsilon) \log \alpha$, where $\alpha$ is the desired cover constraint.
Summary: Monotone Submodular Maximization

- Only makes sense when there is a constraint.
Summary: Monotone Submodular Maximization

- Only makes sense when there is a constraint.
- We discussed cardinality constraint
Only makes sense when there is a constraint.

We discussed cardinality constraint

Generalizes the max $k$-cover problem, and also similar to the set cover problem.
Summary: Monotone Submodular Maximization

- Only makes sense when there is a constraint.
- We discussed cardinality constraint
- Generalizes the max $k$-cover problem, and also similar to the set cover problem.
- Simple greedy algorithm gets $1 - e^{-\ell/k}$ approximation, where $k$ is size of optimal set we compare against, and $\ell$ is size of set greedy algorithm chooses.
Summary: Monotone Submodular Maximization

- Only makes sense when there is a constraint.
- We discussed cardinality constraint
- Generalizes the max $k$-cover problem, and also similar to the set cover problem.
- Simple greedy algorithm gets $1 - e^{-\ell/k}$ approximation, where $k$ is size of optimal set we compare against, and $\ell$ is size of set greedy algorithm chooses.
- Submodular cover: min. $|S|$ s.t. $f(S) \geq \alpha$. 
Only makes sense when there is a constraint.

We discussed cardinality constraint.

Generalizes the max $k$-cover problem, and also similar to the set cover problem.

Simple greedy algorithm gets $1 - e^{-\ell/k}$ approximation, where $k$ is size of optimal set we compare against, and $\ell$ is size of set greedy algorithm chooses.

Submodular cover: min. $|S|$ s.t. $f(S) \geq \alpha$.

Minoux’s accelerated greedy trick.