Submodular Functions, Optimization, and Applications to Machine Learning
— Spring Quarter, Lecture 18 —

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

Prof. Jeff Bilmes

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

May 30th, 2018

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

Announcements, Assignments, and Reminders

- Take home final exam (like long homework). Due Friday, June 8th, 4:00pm via our assignment dropbox (https://canvas.uw.edu/courses/1216339/assignments).
- Get started now. At least read through everything and ask any questions you might have.
- As always, if you have any questions about anything, please ask them via our discussion board (https://canvas.uw.edu/courses/1216339/discussion_topics). Can meet at odd hours via zoom (send message on canvas to schedule time to chat).
Most violated inequality problem in matroid polytope case

Consider

$$P_r^+ = \{ x \in \mathbb{R}^E : x \geq 0, x(A) \leq r_M(A), \forall A \subseteq E \}$$  \hspace{1cm} (18.22)

Suppose we have any $x \in \mathbb{R}^E_+$ such that $x \notin P_r^+$.

Hence, there must be a set of $\mathcal{W} \subseteq 2^V$, each member of which corresponds to a violated inequality, i.e., equations of the form $x(A) > r_M(A)$ for $A \in \mathcal{W}$.

The most violated inequality when $x$ is considered w.r.t. $P_r^+$ corresponds to the set $A$ that maximizes $x(A) - r_M(A)$, i.e., the most violated inequality is valued as:

$$\max \{ x(A) - r_M(A) : A \in \mathcal{W} \} = \max \{ x(A) - r_M(A) : A \subseteq E \}$$  \hspace{1cm} (18.23)

Since $x$ is modular and $x(E \setminus A) = x(E) - x(A)$, we can express this via a min as in:

$$\min \{ r_M(A) + x(E \setminus A) : A \subseteq E \}$$  \hspace{1cm} (18.24)
Most violated inequality/polymatroid membership/SFM

- Consider

\[ P_f^+ = \{ x \in \mathbb{R}^E : x \geq 0, x(A) \leq f(A), \forall A \subseteq E \} \]  \hspace{1cm} (18.22)

- Suppose we have any \( x \in \mathbb{R}^E_+ \) such that \( x \notin P_f^+ \).

- Hence, there must be a set of \( \mathcal{W} \subseteq 2^V \), each member of which corresponds to a violated inequality, i.e., equations of the form \( x(A) > r_M(A) \) for \( A \in \mathcal{W} \).

\[ \mathcal{W} = \{ \{1\}, \{2\} \} \]  \hspace{1cm} \[ \mathcal{W} = \{ \{2\}, \{1, 2\} \} \]  \hspace{1cm} \[ \mathcal{W} = \{ \{1, 2\} \} \]

- The most violated inequality when \( x \) is considered w.r.t. \( P_f^+ \) corresponds to the set \( A \) that maximizes \( x(A) - f(A) \), i.e., the most violated inequality is valuated as:

\[ \max \{ x(A) - f(A) : A \in \mathcal{W} \} = \max \{ x(A) - f(A) : A \subseteq E \} \]  \hspace{1cm} (18.22)

- Since \( x \) is modular and \( x(E \setminus A) = x(E) - x(A) \), we can express this via a min as in:

\[ \min \{ f(A) + x(E \setminus A) : A \subseteq E \} \]  \hspace{1cm} (18.23)

- More importantly, \( \min \{ f(A) + x(E \setminus A) : A \subseteq E \} \) is a form of submodular function minimization, namely

\[ \min \{ f(A) - x(A) : A \subseteq E \} \]  for a submodular \( f \) and \( x \in \mathbb{R}^E_+ \), consisting of a difference of polymatroid and modular function (so \( f - x \) is no longer necessarily monotone, nor positive).

- We will ultimatley answer how general this form of SFM is.
Fundamental circuits in matroids

Lemma 18.2.5

Let \( I \in \mathcal{I}(M) \), and \( e \in E \), then \( I \cup \{e\} \) contains at most one circuit in \( M \).

Proof.

- Suppose, to the contrary, that there are two distinct circuits \( C_1, C_2 \) such that \( C_1 \cup C_2 \subseteq I \cup \{e\} \).
- Then \( e \in C_1 \cap C_2 \), and by (C2), there is a circuit \( C_3 \) of \( M \) s.t. \( C_3 \subseteq (C_1 \cup C_2) \setminus \{e\} \subseteq I \)
- This contradicts the independence of \( I \).

In general, let \( C(I, e) \) be the unique circuit associated with \( I \cup \{e\} \) (commonly called the fundamental circuit in \( M \) w.r.t. \( I \) and \( e \)).

Matroids: The Fundamental Circuit

- Define \( C(I, e) \) be the unique circuit associated with \( I \cup \{e\} \) (the fundamental circuit in \( M \) w.r.t. \( I \) and \( e \), if it exists).
- If \( e \in \text{span}(I) \setminus I \), then \( C(I, e) \) is well defined (\( I + e \) creates one circuit).
- If \( e \in I \), then \( I + e = I \) doesn’t create a circuit. In such cases, \( C(I, e) \) is not really defined.
- In such cases, we define \( C(I, e) = \{e\} \), and we will soon see why.
- If \( e \notin \text{span}(I) \) (i.e., when \( I + e \) is independent), then we set \( C(I, e) = \emptyset \).
The sat function $= \text{Polymatroid Closure}$

- Thus, in a matroid, closure (span) of a set $A$ are all items that $A$ spans (eq. that depend on $A$).
- We wish to generalize closure to polymatroids.
- Consider $x \in P_f$ for polymatroid function $f$.
- Again, recall, tight sets are closed under union and intersection, and therefore form a distributive lattice.
- That is, we saw in Lecture 7 that for any $A, B \in D(x)$, we have that $A \cup B \in D(x)$ and $A \cap B \in D(x)$, which can constitute a join and meet.
- Recall, for a given $x \in P_f$, we have defined this tight family as
  \[ D(x) = \{ A : A \subseteq E, x(A) = f(A) \} \] (18.23)

Minimizers of a Submodular Function form a lattice

Theorem 18.2.6

For arbitrary submodular $f$, the minimizers are closed under union and intersection. That is, let $M = \arg\min_{X \subseteq E} f(X)$ be the set of minimizers of $f$. Let $A, B \in M$. Then $A \cup B \in M$ and $A \cap B \in M$.

Proof.

Since $A$ and $B$ are minimizers, we have $f(A) = f(B) \leq f(A \cap B)$ and $f(A) = f(B) \leq f(A \cup B)$.

By submodularity, we have

\[ f(A) + f(B) \geq f(A \cup B) + f(A \cap B) \] (18.25)

Hence, we must have $f(A) = f(B) = f(A \cup B) = f(A \cap B)$.

Thus, the minimizers of a submodular function form a lattice, and there is a maximal and a minimal minimizer of every submodular function.
The sat function $\equiv$ Polymatroid Closure

- Matroid closure is generalized by the unique maximal element in $D(x)$, also called the polymatroid closure or sat (saturation function).
- For some $x \in P_f$, we have defined:

$$
\begin{align*}
\text{cl}(x) \overset{\text{def}}{=} & \text{sat}(x) \overset{\text{def}}{=} \bigcup \{A : A \in D(x)\} \\
= & \bigcup \{A : A \subseteq E, x(A) = f(A)\} \\
= & \{e : e \in E, \forall \alpha > 0, x + \alpha 1_e \notin P_f\}
\end{align*}
$$

(18.25)

(18.26)

(18.27)

- Hence, sat$(x)$ is the maximal (zero-valued) minimizer of the submodular function $f_x(A) \triangleq f(A) - x(A)$.
- Eq. (18.27) says that sat consists of elements of E for point $x$ that are $P_f$ saturated (any additional positive movement, in that dimension, leaves $P_f$). We’ll revisit this in a few slides.
- First, we see how sat generalizes matroid closure.

Lemma 18.2.6 (Matroid sat : $\mathbb{R}_+^E \rightarrow 2^E$ is the same as closure.)

For $I \in I$, we have $\text{sat}(1_I) = \text{span}(I)$

(18.29)

Proof.

- For $1_I(I) = |I| = r(I)$, so $I \in D(1_I)$ and $I \subseteq \text{sat}(1_I)$. Also, $I \subseteq \text{span}(I)$.
- Consider some $b \in \text{span}(I) \setminus I$.
- Then $I \cup \{b\} \in D(1_I)$ since $1_I(I \cup \{b\}) = |I| = r(I \cup \{b\}) = r(I)$.
- Thus, $b \in \text{sat}(1_I)$.
- Therefore, $\text{sat}(1_I) \supseteq \text{span}(I)$.

...
The sat function, span, and submodular function minimization

- Thus, for a matroid, sat(1_I) is exactly the closure (or span) of I in the matroid. I.e., for matroid (E, r), we have span(I) = sat(1_B).
- Recall, for x ∈ P_f and polymatroidal f, sat(x) is the maximal (by inclusion) minimizer of f(A) − x(A), and thus in a matroid, span(I) is the maximal minimizer of the submodular function formed by r(A) − 1_I(A).
- Submodular function minimization can solve “span” queries in a matroid or “sat” queries in a polymatroid.

sat, as tight polymatroidal elements

- We are given an x ∈ P_f^+ for submodular function f.
- Recall that for such an x, sat(x) is defined as

  \[ sat(x) = \bigcup \{ A : x(A) = f(A) \} \]  

  (18.1)

- We also have stated that sat(x) can be defined as:

  \[ sat(x) = \left\{ e : \forall \alpha > 0, x + \alpha 1_e \notin P_f^+ \right\} \]  

  (18.2)

- We next show more formally that these are the same.
sat, as tight polymatroidal elements

- Let's start with one definition and derive the other.
  \[
  \text{sat}(x) \overset{\text{def}}{=} \left\{ e : \forall \alpha > 0, x + \alpha 1_e \notin P^+_f \right\} \quad (18.3)
  \]
  \[
  = \left\{ e : \forall \alpha > 0, \exists A \text{ s.t. } (x + \alpha 1_e)(A) > f(A) \right\} \quad (18.4)
  \]
  \[
  = \left\{ e : \forall \alpha > 0, \exists A \ni e \text{ s.t. } (x + \alpha 1_e)(A) > f(A) \right\} \quad (18.5)
  \]
  - This last bit follows since \( 1_e(A) = 1 \iff e \in A \). Continuing, we get
  \[
  \text{sat}(x) = \left\{ e : \forall \alpha > 0, \exists A \ni e \text{ s.t. } (x + \alpha 1_e)(A) + \alpha > f(A) \right\} \quad (18.6)
  \]
  - Given that \( x \in P^+_f \), meaning \( x(A) \leq f(A) \) for all \( A \), we must have
  \[
  \text{sat}(x) = \left\{ e : \forall \alpha > 0, \exists A \ni e \text{ s.t. } x(A) + \alpha = f(A) \right\} \quad (18.7)
  \]
  \[
  = \left\{ e : \exists A \ni e \text{ s.t. } x(A) = f(A) \right\} \quad (18.8)
  \]
  - So now, if \( A \) is any set such that \( x(A) = f(A) \), then we clearly have
  \[
  \forall e \in A, e \in \text{sat}(x), \text{ and therefore that } \text{sat}(x) \supseteq A \quad (18.9)
  \]

- And therefore, with sat as defined in Eq. (17.35),
  \[
  \text{sat}(x) \supseteq \bigcup \{ A : x(A) = f(A) \} \quad (18.10)
  \]
  - On the other hand, for any \( e \in \text{sat}(x) \) defined as in Eq. (18.8), since \( e \) is itself a member of a tight set, there is a set \( A \ni e \) such that \( x(A) = f(A) \), giving
  \[
  \text{sat}(x) \subseteq \bigcup \{ A : x(A) = f(A) \} \quad (18.11)
  \]
  - Therefore, the two definitions of sat are identical.
Another useful concept is saturation capacity which we develop next.

For \( x \in P_f \), and \( e \in E \), consider finding
\[
\max \{ \alpha : \alpha \in \mathbb{R}, x + \alpha 1_e \in P_f \} \quad (18.12)
\]

This is identical to:
\[
\max \{ \alpha : (x + \alpha 1_e)(A) \leq f(A), \forall A \supseteq \{e\} \} \quad (18.13)
\]
since any \( B \subseteq E \) such that \( e \notin B \) does not change in a \( 1_e \) adjustment, meaning \( (x + \alpha 1_e)(B) = x(B) \).

Again, this is identical to:
\[
\max \{ \alpha : x(A) + \alpha \leq f(A), \forall A \supseteq \{e\} \} \quad (18.14)
\]
or
\[
\max \{ \alpha : \alpha \leq f(A) - x(A), \forall A \supseteq \{e\} \} \quad (18.15)
\]

The max is achieved when
\[
\alpha = \hat{c}(x; e) \overset{\text{def}}{=} \min \{ f(A) - x(A), \forall A \supseteq \{e\} \} \quad (18.16)
\]
\( \hat{c}(x; e) \) is known as the saturation capacity associated with \( x \in P_f \) and \( e \).

Thus we have for \( x \in P_f \),
\[
\hat{c}(x; e) \overset{\text{def}}{=} \min \{ f(A) - x(A), \forall A \ni \{e\} \} = \max \{ \alpha : \alpha \in \mathbb{R}, x + \alpha 1_e \in P_f \} \quad (18.17)
\]
We immediately see that for \( e \in E \setminus \text{sat}(x) \), we have that \( \hat{c}(x; e) > 0 \).
Also, we have that: \( e \in \text{sat}(x) \iff \hat{c}(x; e) = 0 \).
Note that any \( \alpha \) with \( 0 \leq \alpha \leq \hat{c}(x; e) \) we have \( x + \alpha 1_e \in P_f \).
We also see that computing \( \hat{c}(x; e) \) is a form of submodular function minimization.
Dependence Function

- Tight sets can be restricted to contain a particular element.
- Given \( x \in P_f \), and \( e \in \text{sat}(x) \), define
  \[
  D(x, e) = \{ A : e \in A \subseteq E, x(A) = f(A) \} 
  \]
  \[
  = D(x) \cap \{ A : A \subseteq E, e \in A \} 
  \]
  (18.19)
  (18.20)
- Thus, \( D(x, e) \subseteq D(x) \), and \( D(x, e) \) is a sublattice of \( D(x) \).
- Therefore, we can define a unique minimal element of \( D(x, e) \) denoted as follows:
  \[
  \text{dep}(x, e) = \begin{cases} 
  \bigcap \{ A : e \in A \subseteq E, x(A) = f(A) \} & \text{if } e \in \text{sat}(x) \\
  \emptyset & \text{else}
  \end{cases} 
  \]
  (18.21)
- i.e., \( \text{dep}(x, e) \) is the minimal element in \( D(x) \) that contains \( e \) (the minimal \( x \)-tight set containing \( e \)).
dep and sat in a lattice

- Given $x \in P_f$, recall distributive lattice of tight sets $D(x) = \{A : x(A) = f(A)\}$
- We had that $\text{sat}(x) = \bigcup \{A : A \in D(x)\}$ is the “1” element of this lattice.
- Consider the “0” element of $D(x)$, i.e., $\text{dry}(x) \overset{\text{def}}{=} \bigcap \{A : A \in D(x)\}$
- We can see $\text{dry}(x)$ as the elements that are necessary for tightness.
- That is, we can equivalently define $\text{dry}(x)$ as
  \[
  \text{dry}(x) = \{e' : x(A) < f(A), \forall A \not\ni e'\} \tag{18.22}
  \]
- This can be read as, for any $e' \in \text{dry}(x)$, any set that does not contain $e'$ is not tight for $x$ (any set $A$ that is missing any element of $\text{dry}(x)$ is not tight).
- Perhaps, then, a better name for $\text{dry}$ is $\text{ntight}(x)$, for the necessary for tightness (but we’ll actually use neither name).
- Note that $\text{dry}$ need not be the empty set. Exercise: give example.

An alternate expression for $\text{dep} = \text{dry}$: restated

- Now, given $x \in P_f$, and $e \in \text{sat}(x)$, recall distributive sub-lattice of $e$-containing tight sets $D(x, e) = \{A : e \in A, x(A) = f(A)\}$
- We can define the “1” element of this sub-lattice as $\text{sat}(x, e) \overset{\text{def}}{=} \bigcup \{A : A \in D(x, e)\}$.
- Analogously, we can define the “0” element of this sub-lattice as $\text{dry}(x, e) \overset{\text{def}}{=} \bigcap \{A : A \in D(x, e)\}$.
- We can see $\text{dry}(x, e)$ as the elements that are necessary for $e$-containing tightness, with $e \in \text{sat}(x)$.
- That is, we can view $\text{dry}(x, e)$ as
  \[
  \text{dry}(x, e) = \{e' : x(A) < f(A), \forall A \not\ni e', e \in A\} \tag{18.23}
  \]
- This can be read as, for any $e' \in \text{dry}(x, e)$, any $e$-containing set that does not contain $e'$ is not tight for $x$.
- But actually, $\text{dry}(x, e) = \text{dep}(x, e)$, so we have derived another expression for $\text{dep}(x, e)$ in Eq. (18.23).
Now, let \((E, \mathcal{I}) = (E, r)\) be a matroid, and let \(I \in \mathcal{I}\) giving \(1_I \in P_r\). We have \(\text{sat}(1_I) = \text{span}(I) = \text{closure}(I)\).

- Given \(e \in \text{sat}(1_I) \setminus I\) and then consider an \(A \ni e\) with \(|I \cap A| = r(A)\).
- Then \(I \cap A\) serves as a base for \(A\) (i.e., \(I \cap A\) spans \(A\)) and any such \(A\) contains a circuit (i.e., we can add \(e \in A \setminus I\) to \(I \cap A\) w/o increasing rank).

- Given \(e \in \text{sat}(1_I) \setminus I\), and consider \(\text{dep}(1_I, e)\), with
  \[
  \text{dep}(1_I, e) = \bigcap \{A : e \in A \subseteq E, 1_I(A) = r(A)\} \tag{18.24}
  \]
  \[
  = \bigcap \{A : e \in A \subseteq E, |I \cap A| = r(A)\} \tag{18.25}
  \]
  \[
  = \bigcap \{A : e \in A \subseteq E, r(A) - |I \cap A| = 0\} \tag{18.26}
  \]

By SFM lattice, \(\exists\) a unique minimal \(A \ni e\) with \(|I \cap A| = r(A)\).

Thus, \(\text{dep}(1_I, e)\) must be a circuit since if it included more than a circuit, it would not be minimal in this sense.

Therefore, when \(e \in \text{sat}(1_I) \setminus I\), then \(\text{dep}(1_I, e) = C(I, e)\) where \(C(I, e)\) is the unique circuit contained in \(I + e\) in a matroid (the fundamental circuit of \(e\) and \(I\) that we encountered before).

- Now, if \(e \in \text{sat}(1_I) \cap I\) with \(I \in \mathcal{I}\), we said that \(C(I, e)\) was undefined (since no circuit is created in this case) and so we defined it as \(C(I, e) = \{e\}\).
- In this case, for such an \(e\), we have \(\text{dep}(1_I, e) = \{e\}\) since all such sets \(A \ni e\) with \(|I \cap A| = r(A)\) contain \(e\), but in this case no cycle is created, i.e., \(|I \cap A| \geq |I \cap \{e\}| = r(e) = 1\).
- We are thus free to take subsets of \(I\) as \(A\), all of which must contain \(e\), but all of which have rank equal to size, and min size is 1.
- Also note: in general for \(x \in P_f\) and \(e \in \text{sat}(x)\), we have \(\text{dep}(x, e)\) is tight by definition (i.e., \(x(\text{dep}(x, e)) = f(\text{dep}(x, e))\)).
Summary of sat, and dep

- For $x \in P_f$, $\text{sat}(x)$ (span, closure) is the maximal saturated ($x$-tight) set w.r.t. $x$. I.e., $\text{sat}(x) = \{ e : e \in E, \forall \alpha > 0, x + \alpha e \notin P_f \}$. That is,

\[
\text{cl}(x) \overset{\text{def}}{=} \text{sat}(x) = \bigcup \{ A : A \in D(x) \} \tag{18.27}
\]

\[
= \bigcup \{ A : A \subseteq E, x(A) = f(A) \} \tag{18.28}
\]

\[
= \{ e : e \in E, \forall \alpha > 0, x + \alpha e \notin P_f \} \tag{18.29}
\]

- For $e \in \text{sat}(x)$, we have $\text{dep}(x, e) \subseteq \text{sat}(x)$ (fundamental circuit) is the minimal (common) saturated ($x$-tight) set w.r.t. $x$ containing $e$. I.e.,

\[
\text{dep}(x, e) = \begin{cases} 
\bigcap \{ A : e \in A \subseteq E, x(A) = f(A) \} & \text{if } e \in \text{sat}(x) \\
\emptyset & \text{else}
\end{cases}
\]

\[
= \{ e' : \exists \alpha > 0, \text{ s.t. } x + \alpha(1_e - 1_{e'}) \in P_f \} \tag{18.30}
\]

Note, if $x + \alpha(1_e - 1_{e'}) \in P_f$, then $x + \alpha'(1_e - 1_{e'}) \in P_f$ for any $0 \leq \alpha' < \alpha$.

Dependence Function and exchange

- For $e \in \text{span}(I) \setminus I$, we have that $I + e \notin I$. This is a set addition restriction property.
- Analogously, for $e \in \text{sat}(x)$, any $x + \alpha 1_e \notin P_f$ for $\alpha > 0$. This is a vector increase restriction property.
- Recall, we have $C(I, e) \setminus e' \in I$ for $e' \in C(I, e)$. I.e., $C(I, e)$ consists of elements that when removed recover independence.
- In other words, for $e \in \text{span}(I) \setminus I$, we have that

\[
C(I, e) = \{ a \in E : I + e - a \in I \} \tag{18.31}
\]

I.e., an addition of $e$ to $I$ stays within $I$ only if we simultaneously remove one of the elements of $C(I, e)$.
- But, analogous to the circuit case, is there an exchange property for $\text{dep}(x, e)$ in the form of vector movement restriction?
- We might expect the vector $\text{dep}(x, e)$ property to take the form: a positive move in the $e$-direction stays within $P_f^+$ only if we simultaneously take a negative move in one of the $\text{dep}(x, e)$ directions.
Dependence Function and exchange in 2D

- \( \text{dep}(x, e) \) is set of negative directions we must move if we want to move in positive direction, starting at \( x \) and staying within \( P_f \).

- Viewable in 2D, we have for \( A, B \subseteq E, A \cap B = \emptyset \):

  \[
  \text{Left: } e \in B \text{ and } A \cap \text{dep}(x, e) = \emptyset, \text{ and we can’t move further in (e) direction, and moving in any negative } a \in A \text{ direction doesn’t change that. No dependence between (e) and any element in } A.
  \]

  \[
  \text{Right: } A \subseteq \text{dep}(x, e). \text{ We can’t move further in the (e) direction, but we can move further in (e) direction by moving in some negative } a \in A \text{ direction. Dependence between (e) and elements in } A.
  \]

Dependence Function and exchange in 3D

- We can move neither in the (e) nor the (a) direction, but we can move in the (e) direction if we simultaneously move in the -(a) direction.

- In 3D, we have:

  \[
  \text{l.e., for } e \in \text{sat}(x), a \in \text{sat}(x), a \in \text{dep}(x, e), e \notin \text{dep}(x, a), \text{ and } \text{dep}(x, e) = \{ a : a \in E, \exists \alpha > 0 : x + \alpha(1_e - 1_a) \in P_f \} \quad \text{(18.32)}
  \]

- We next show this formally.
The derivation for \( \text{dep}(x, e) \) involves turning a strict inequality into a non-strict one with a strict explicit slack variable \( \alpha \):

\[
\text{dep}(x, e) = \text{ntight}(x, e) = \{ e' : x(A) < f(A), \forall A \not\ni e' \} \quad (18.33)
\]

\[
= \{ e' : \exists \alpha > 0, \text{ s.t. } \alpha \leq f(A) - x(A), \forall A \not\ni e' \} \quad (18.34)
\]

\[
= \{ e' : \exists \alpha > 0, \text{ s.t. } \alpha 1_e(A) \leq f(A) - x(A), \forall A \not\ni e' \} \quad (18.35)
\]

\[
= \{ e' : \exists \alpha > 0, \text{ s.t. } (1_e(A) - 1_{e'}(A)) \leq f(A) - x(A), \forall A \not\ni e', e \in A \} \quad (18.36)
\]

\[
= \{ e' : \exists \alpha > 0, \text{ s.t. } x(A) + \alpha (1_e(A) - 1_{e'}(A)) \leq f(A), \forall A \not\ni e', e \in A \} \quad (18.37)
\]

Now, \( 1_e(A) - 1_{e'}(A) = 0 \) if either \( \{ e, e' \} \subseteq A \), or \( \{ e, e' \} \cap A = \emptyset \).

Also, if \( e' \in A \) but \( e \not\in A \), then

\[
x(A) + \alpha (1_e(A) - 1_{e'}(A)) = x(A) - \alpha \leq f(A) \text{ since } x \in P_f.
\]

thus, we get the same in the above if we remove the constraint \( A \not\ni e', e \in A \), that is we get

\[
\text{dep}(x, e) = \{ e' : \exists \alpha > 0, \text{ s.t. } x(A) + \alpha (1_e(A) - 1_{e'}(A)) \leq f(A), \forall A \} \quad (18.39)
\]

This is then identical to

\[
\text{dep}(x, e) = \{ e' : \exists \alpha > 0, \text{ s.t. } x + \alpha (1_e - 1_{e'}) \in P_f \} \quad (18.40)
\]

Compare with original, the minimal element of \( D(x, e) \), with \( e \in \text{sat}(x) \):

\[
\text{dep}(x, e) = \begin{cases} 
\bigcap \{ A : e \in A \subseteq E, x(A) = f(A) \} & \text{if } e \in \text{sat}(x) \\
\emptyset & \text{else}
\end{cases} \quad (18.41)
\]
Summary of Concepts

- Most violated inequality $\max \{x(A) - f(A) : A \subseteq E\}$
- Matroid by circuits, and the fundamental circuit $C(I, e) \subseteq I + e$.
- Minimizers of submodular functions form a lattice.
- Minimal and maximal element of a lattice.
- $x$-tight sets, maximal and minimal tight set.
- $\text{sat}$ function & Closure
- Saturation Capacity
- $\epsilon$-containing tight sets
- dep function & fundamental circuit of a matroid

Summary important definitions so far: tight, dep, & sat

- $x$-tight sets: For $x \in P_f$, $D(x) \triangleq \{A \subseteq E : x(A) = f(A)\}$.
- Polymatroid closure/maximal $x$-tight set: For $x \in P_f$,
  $\text{sat}(x) \triangleq \cup\{A : A \in D(x)\} = \{e : e \in E, \forall \alpha > 0, x + \alpha 1_e \notin P_f\}$.
- Saturation capacity: for $x \in P_f$, $0 \leq \hat{c}(x; e) \triangleq \min \{f(A) - x(A) : \forall A \ni e\} = \max \{\alpha : \alpha \in \mathbb{R}, x + \alpha 1_e \in P_f\}$
- Recall: $\text{sat}(x) = \{e : \hat{c}(x; e) = 0\}$ and $E \setminus \text{sat}(x) = \{e : \hat{c}(x; e) > 0\}$.
- $\epsilon$-containing $x$-tight sets: For $x \in P_f$,
  $D(x, e) = \{A : e \in A \subseteq E, x(A) = f(A)\} \subseteq D(x)$.
- Minimal $\epsilon$-containing $x$-tight set/polymatroidal fundamental circuit\/: For $x \in P_f$,
  $\text{dep}(x, e) = \begin{cases} \cap \{A : e \in A \subseteq E, x(A) = f(A)\} & \text{if } e \in \text{sat}(x) \\ \emptyset & \text{else} \end{cases}$
  $= \{e' : \exists \alpha > 0, \text{ s.t. } x + \alpha (1_e - 1_{e'}) \in P_f\}$
Submodular Function Minimization (SFM) and Min-Norm

- We saw that SFM can be used to solve most violated inequality problems for a given \( x \in P_f \) and, in general, SFM can solve the question “Is \( x \in P_f \)” by seeing if \( x \) violates any inequality (if the most violated one is negative, solution to SFM, then \( x \in P_f \)).

- Unconstrained SFM, \( \min_{A \subseteq V} f(A) \) solves many other problems as well in combinatorial optimization, machine learning, and other fields.

- We next study an algorithm, the “Fujishige-Wolfe Algorithm”, or what is known as the “Minimum Norm Point” algorithm, which is an active set method to do this, and one that in practice works about as well as anything else people (so far) have tried for general purpose SFM.

- Note special case SFM can be much faster.

Min-Norm Point: Definition

- Consider the optimization:

\[
\begin{align*}
\text{minimize} & \quad \|x\|_2^2 \\
\text{subject to} & \quad x \in B_f
\end{align*}
\]

where \( B_f \) is the base polytope of submodular \( f \), and \( \|x\|_2^2 = \sum_{e \in E} x(e)^2 \) is the squared 2-norm. Let \( x^* \) be the optimal solution.

- Note, \( x^* \) is the unique optimal solution since we have a strictly convex objective over a set of convex constraints.

- \( x^* \) is called the minimum norm point of the base polytope.
Consider submodular function $f : 2^V \to \mathbb{R}$ with $|V| = 4$, and for $X \subseteq V$, concave $g$,

$$f(X) = g(|X|) = \sum_{i=1}^{|X|} (4 - i + 1)$$

Then $B_f$ is a 3D polytope, and in this particular case gives us a permutahedron with 24 distinct extreme points, on the right (from wikipedia).
Min-Norm Point and Submodular Function Minimization

- Given optimal solution $x^*$ to the above, consider the quantities

$$y^* = x^* \land 0 = (\min(x^*(e), 0) | e \in E) \quad (18.43)$$

$$A_- = \{ e : x^*(e) < 0 \} \quad (18.44)$$

$$A_0 = \{ e : x^*(e) \leq 0 \} \quad (18.45)$$

- Thus, we immediately have that:

$$A_- \subseteq A_0 \quad (18.46)$$

and that

$$x^*(A_-) = x^*(A_0) = y^*(A_-) = y^*(A_0) \quad (18.47)$$

- It turns out, these quantities will solve the submodular function minimization problem, as we now show.

- The proof is nice since it uses the tools we’ve been recently developing.

More about the base $B_f$

**Theorem 18.6.1**

Let $f$ be a polymatroid function and suppose that $E$ can be partitioned into $(E_1, E_2, \ldots, E_k)$ such that $f(A) = \sum_{i=1}^k f(A \cap E_i)$ for all $A \subseteq E$, and $k$ is maximum. Then the base polytope $B_f = \{ x \in P_f : x(E) = f(E) \}$ (the $E$-tight subset of $P_f$) has dimension $|E| - k$.

- In fact, every $x \in P_f$ is dominated by $x \leq y \in B_f$.

**Theorem 18.6.2**

If $x \in P_f$ and $T$ is tight for $x$ (meaning $x(T) = f(T)$), then there exists $y \in B_f$ with $x \leq y$ and $y(e) = x(e)$ for $e \in T$.

- We will prove these after we describe min-norm algorithm.
The following slide repeats Theorem 12.3.2 from lecture 12 and is one of the most important theorems in submodular theory.

**Theorem 18.6.1**

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

$$
\text{rank}(x) = \max \{y(E) : y \leq x, y \in P_f\} = \min \{x(A) + f(E \setminus A) : A \subseteq E\}
$$  \hspace{1cm} (18.1)

Essentially the same theorem as Theorem 10.4.1, but note $P_f$ rather than $P_f^+$. Taking $x = 0$ we get:

**Corollary 18.6.2**

Let $f$ be a submodular function defined on subsets of $E$. We have:

$$
\text{rank}(0) = \max \{y(E) : y \leq 0, y \in P_f\} = \min \{f(A) : A \subseteq E\}
$$  \hspace{1cm} (18.2)
Modified max-min theorem

- Min-max theorem (Thm 12.3.2) restated for $x = 0$.

\[
\max \{ y(E) | y \in P_f, y \leq 0 \} = \min \{ f(X) | X \subseteq V \} \quad (18.48)
\]

**Theorem 18.6.3 (Edmonds-1970)**

\[
\min \{ f(X) | X \subseteq E \} = \max \{ x^{-}(E) | x \in B_f \} \quad (18.49)
\]

where $x^{-}(e) = \min \{ x(e), 0 \}$ for $e \in E$.

**Proof via the Lovász ext.**

\[
\min \{ f(X) | X \subseteq E \} = \min_{w \in [0,1]^E} \tilde{f}(w) = \min_{w \in [0,1]^E} \max_{x \in P_f} w^T x \quad (18.50)
\]

\[
= \min_{w \in [0,1]^E} \max_{x \in B_f} w^T x \quad (18.51)
\]

\[
= \max_{x \in B_f} \min_{w \in [0,1]^E} w^T x \quad (18.52)
\]

\[
= \max_{x \in B_f} x^{-}(E) \quad (18.53)
\]

Convexity, Strong duality, and min/max swap

The min/max switch follows from strong duality. I.e., consider $g(w, x) = w^T x$ and we have domains $w \in [0,1]^E$ and $x \in B_f$. then for any $(w, x) \in [0,1]^E \times B_f$, we have

\[
\min_{w' \in [0,1]^E} g(w', x) \leq g(w, x) \leq \max_{x' \in B_f} g(w, x') \quad (18.54)
\]

which means that we have weak duality

\[
\max_{x \in B_f} \min_{w' \in [0,1]^E} g(w', x) \leq \min_{w \in [0,1]^E} \max_{x' \in B_f} g(w, x') \quad (18.55)
\]

but since $g(w, x)$ is linear, we have strong duality, meaning

\[
\max_{x \in B_f} \min_{w' \in [0,1]^E} g(w', x) = \min_{w \in [0,1]^E} \max_{x' \in B_f} g(w, x') \quad (18.56)
\]
Alternate proof of modified max-min theorem

We start directly from Theorem 12.3.2.

\[
\max (y(E) : y \leq 0, y \in P_f) = \min (f(A) : A \subseteq E)
\]

(18.57)

Given \(y \in \mathbb{R}^E\), define \(y^{-} \in \mathbb{R}^E\) with \(y^{-}(e) = \min \{y(e), 0\}\) for \(e \in E\).

\[
\max (y(E) : y \leq 0, y \in P_f) = \max (y^{-}(E) : y \leq 0, y \in P_f)
\]

(18.58)

\[
= \max (y^{-}(E) : y \in P_f)
\]

(18.59)

\[
= \max (y^{-}(E) : y \in B_f)
\]

(18.60)

The first equality follows since \(y \leq 0\). For the second equality will be shown on the following slide. The third equality follows since for any \(x \in P_f\) there exists a \(y \in B_f\) with \(x \leq y\) (follows from Theorem 18.6.2).

Alternate proof of modified max-min theorem

Consider the following two problems:

\[
\max \sum_{e \in E} y(e) \quad \text{s.t.} \quad y \leq x \quad y \in P
\]

(18.61a)

\[
\text{max} \sum_{e \in E} \min(y(e), x(e)) \quad \text{s.t.} \quad y \in P
\]

(18.62a)

- Solutions identical cost. Let \(y_1^*\) be l.h.s. OPT and \(y_2^*\) be r.h.s. OPT.
- Consider \(y_1^*\) as r.h.s. solution and suppose it is worse than r.h.s. OPT:

\[
\sum_{e \in E} \min(y_1^*(e), x(e)) < \sum_{e \in E} \min(y_2^*(e), x(e))
\]

(18.63)

Hence, \(\exists e'\) s.t. \(y_1^*(e') < \min(y_2^*(e'), x(e'))\). Recall \(y_1^*, y_2^* \in P\).

- This implies \(\sum_{e \neq e'} y_1^*(e) + y_1^*(e') < \sum_{e \neq e'} y_2^*(e) + \min(y_2^*(e'), x(e'))\), better feasible solution to l.h.s., contradicting \(y_1^*\)'s optimality for l.h.s.
- Similarly, consider \(y_2^*\) as l.h.s. solution, suppose worse than l.h.s. OPT

\[
\sum_{e \in E} y_2^*(e) < \sum_{e \in E} y_1^*(e)
\]

(18.64)

Then \(\exists e'\) such that \(y_2^*(e') < y_1^*(e') \leq x(e')\).

- This implies that replacing \(y_2^*(e')\)'s value with \(y_1^*(e')\) is still feasible for r.h.s. but better, contradicting \(y_2^*\)'s optimality.
Closure/Sat
Fund. Circuit/Dep
Min-Norm Point Definitions
Review & Support for Min-Norm
Proof that min-norm gives optimal
Computing Min-Norm Vector for $B_f$

\[ \min \{ w^\top x : x \in B_f \} \]

- Recall that the greedy algorithm solves, for $w \in \mathbb{R}^E_+$
  \[ \max \{ w^\top x | x \in P_f \} = \max \{ w^\top x | x \in B_f \} \]  \hspace{1cm} (18.66)
  since for all $x \in P_f$, there exists $y \geq x$ with $y \in B_f$.
- For arbitrary $w \in \mathbb{R}^E$, greedy algorithm will also solve:
  \[ \max \{ w^\top x | x \in B_f \} \]  \hspace{1cm} (18.67)
- Also, since $w \in \mathbb{R}^E$ is arbitrary, and since
  \[ \min \{ w^\top x | x \in B_f \} = - \max \{ -w^\top x | x \in B_f \} \]  \hspace{1cm} (18.68)
  the greedy algorithm using ordering $(e_1, e_2, \ldots, e_m)$ such that
  \[ w(e_1) \leq w(e_2) \leq \cdots \leq w(e_m) \]  \hspace{1cm} (18.69)
  will solve l.h.s. of Equation (18.68).

Let $f(A)$ be arbitrary submodular function, and $f(A) = f'(A) - m(A)$ where $f'$ is polymatroidal, and $w \in \mathbb{R}^E$.

\[
\begin{align*}
\max \{ w^\top x | x \in B_f \} &= \max \{ w^\top x | x(A) \leq f(A) \forall A, x(E) = f(E) \} \\
&= \max \{ w^\top x | x(A) \leq f'(A) - m(A) \forall A, x(E) = f'(E) - m(E) \} \\
&= \max \{ w^\top x | x(A) + m(A) \leq f'(A) \forall A, x(E) + m(E) = f'(E) \} \\
&= \max \{ w^\top x + w^\top m | x(A) + m(A) \leq f'(A) \forall A, x(E) + m(E) = f'(E) \} - w^\top m \\
&= \max \{ w^\top y | y \in B_{f'} \} - w^\top m \\
&= w^\top y^* - w^\top m = w^\top (y^* - m)
\end{align*}
\]

where $y = x + m$, so that $x^* = y^* - m$.

So $y^*$ uses greedy algorithm with positive orthant $B_{f'}$. To show, we use Theorem 11.4.1 in Lecture 11, but we don’t require $y \geq 0$, and don’t stop when $w$ goes negative to ensure $y^* \in B_{f'}$. Then when we subtract off $m$ from $y^*$, we get solution to the original problem.
**Theorem 18.7.1**

Let $y^*$, $A_-$, and $A_0$ be as given. Then $y^*$ is a maximizer of the l.h.s. of Eqn. (18.48). Moreover, $A_-$ is the unique minimal minimizer of $f$ and $A_0$ is the unique maximal minimizer of $f$.

**Proof.**

- First note, since $x^* \in B_f$, we have $x^*(E) = f(E)$, meaning $\text{sat}(x^*) = E$. Thus, we can consider any $e \in E$ within $\text{dep}(x^*, e)$.

- Consider any pair $(e, e')$ with $e' \in \text{dep}(x^*, e)$ and $e \in A_-$. Then $x^*(e) < 0$, and $\exists \alpha > 0$ s.t. $x^* + \alpha 1_e - \alpha 1_{e'} \in P_f$.

- We have $x^*(E) = f(E)$ and $x^*$ is minimum in l2 sense. We have $(x^* + \alpha 1_e - \alpha 1_{e'}) \in P_f$, and in fact

$$
(x^* + \alpha 1_e - \alpha 1_{e'})(E) = x^*(E) + \alpha - \alpha = f(E) \tag{18.70}
$$

so $x^* + \alpha 1_e - \alpha 1_{e'} \in B_f$ also.

---

**proof of Thm. 18.7.1 cont.**

- Then $(x^* + \alpha 1_e - \alpha 1_{e'})(E) = x^*(E \setminus \{e, e'\}) + (x^*(e) + \alpha) + (x^*(e') - \alpha) = f(E)$.

  \[
  \begin{align*}
  x^*_\text{new}(e) + x^*_\text{new}(e')
  \end{align*}
  \]

- Minimality of $x^* \in B_f$ in l2 sense requires that, with such an $\alpha > 0$,

  \[
  \left( x^*(e) \right)^2 + \left( x^*(e') \right)^2 < \left( x^*_\text{new}(e) \right)^2 + \left( x^*_\text{new}(e') \right)^2
  \]

- Given that $e \in A_-$, $x^*(e) < 0$. Thus, if $x^*(e') > 0$, we could have

  \[
  (x^*(e) + \alpha)^2 + (x^*(e') - \alpha)^2 < (x^*(e))^2 + (x^*(e'))^2,
  \]

  contradicting the optimality of $x^*$.

- If $x^*(e') = 0$, we would have

  \[
  (x^*(e) + \alpha)^2 + (\alpha)^2 < (x^*(e))^2,
  \]

  for any $0 < \alpha < |x^*(e)|$ (Exercise), again contradicting the optimality of $x^*$.

- Thus, we must have $x^*(e') < 0$ (strict negativity).
Min-Norm Point and SFM

... proof of Thm. 18.7.1 cont.

- Thus, for a pair \((e, e')\) with \(e' \in \text{dep}(x^*, e)\) and \(e \in A_-\), we have \(x(e') < 0\) and hence \(e' \in A_-\).
- Hence, \(\forall e \in A_-\), we have \(\text{dep}(x^*, e) \subseteq A_-\).
- A very similar argument can show that, \(\forall e \in A_0\), we have \(\text{dep}(x^*, e) \subseteq A_0\).
- Also, recall that \(e \in \text{dep}(x^*, e)\).

Therefore, we have \(\bigcup_{e \in A_-} \text{dep}(x^*, e) = A_-\) and \(\bigcup_{e \in A_0} \text{dep}(x^*, e) = A_0\)

\[\text{dep}(x^*, e)\] is minimal tight set containing \(e\), meaning
\[x^*(\text{dep}(x^*, e)) = f(\text{dep}(x^*, e))\], and since tight sets are closed under union, we have that \(A_-\) and \(A_0\) are also tight, meaning:
\[x^*(A_-) = f(A_-) = f(A_0)\]  \[x^*(A_0) = f(A_0)\]
\[x^*(A_-) = x^*(A_0) = y^*(E) = y^*(A_0) + \left[\frac{y^*(E \setminus A_0)}{0}\right]\]  \[= 0\]

and therefore, all together we have
\[f(A_-) = f(A_0) = x^*(A_-) = x^*(A_0) = y^*(E)\]  \[= 0\]

Hence, \(f(A_-) = f(A_0)\), meaning \(A_-\) and \(A_0\) have the same valuation, but we have not yet shown they are the minimizers of the submodular function, nor that they are, resp. the maximal and minimal minimizers.
Min-Norm Point and SFM

...proof of Thm. 18.7.1 cont.

- Now, $y^*$ is feasible for the l.h.s. of Eqn. (18.48) (recall, which is $\max\{y(E)|y \in P_f, y \leq 0\} = \min\{f(X)|X \subseteq V\}$). This follows since, we have $y^* = x^* \wedge 0 \leq 0$, and since $x^* \in B_f \subset P_f$, and $y^* \leq x^*$ and $P_f$ is down-closed, we have that $y^* \in P_f$.

- Also, for any $y \in P_f$ with $y \leq 0$ and for any $X \subseteq E$, we have $y(E) \leq y(X) \leq f(X)$.

- Hence, we have found a feasible for l.h.s. of Eqn. (18.48), $y^* \leq 0$, $y^* \in P_f$, so $y^*(E) \leq f(X)$ for all $X$.

- So $y^*(E) \leq \min\{f(X)|X \subseteq V\}$.

- Considering Eqn. (18.75), we have found sets $A_-$ and $A_0$ with tightness in Eqn. (18.48), meaning $y^*(E) = f(A_-) = f(A_0)$.

- Hence, $y^*$ is a maximizer of l.h.s. of Eqn. (18.48), and $A_-$ and $A_0$ are minimizers of $f$.

We next show that, not only are they minimizers, but $A_-$ is the unique minimal and $A_0$ is the unique maximal minimizer of $f$.

- Now, for any $X \subset A_-$, we have

  \[ f(X) \geq x^*(X) > x^*(A_-) = f(A_-) \quad (18.75) \]

- And for any $X \supset A_0$, we have

  \[ f(X) \geq x^*(X) > x^*(A_0) = f(A_0) \quad (18.76) \]

- Hence, $A_-$ must be the unique minimal minimizer of $f$, and $A_0$ is the unique maximal minimizer of $f$. 
Min-Norm Point and SFM

- So, if we have a procedure to compute the min-norm point computation, we can solve SFM.
- Nice thing about previous proof is that it uses both expressions for $\text{dep}$ for different purposes.
- This was discovered by Fujishige (in fact the proof above is an expanded version of the one found in the book).
- As we saw last time, the algorithm (by F. Wolfe) can find this min-norm point, essentially an active-set procedure for quadratic programming. It uses Edmonds's greedy algorithm to make it efficient.
- This is currently the best practical algorithm for general purpose submodular function minimization.
- But recall, its underlying lower-bound complexity is unknown.

Recall, that the set of minimizers of $f$ forms a lattice.

Q: If we take any $A$ with $A_- \subset A \subset A_0$, is $A$ also a minimizer?

In fact, with $x^*$ the min-norm point, and $A_-$ and $A_0$ as defined above, we have the following theorem:

**Theorem 18.7.2**

Let $A \subseteq E$ be any minimizer of submodular $f$, and let $x^*$ be the minimum-norm point. Then $A$ can be expressed in the form:

$$A = A_- \cup \bigcup_{a \in A_m} \text{dep}(x^*, a)$$

(18.77)

for some set $A_m \subseteq A_0 \setminus A_-$. Conversely, for any set $A_m \subseteq A_0 \setminus A_-$, then $A \triangleq A_- \cup \bigcup_{a \in A_m} \text{dep}(x^*, a)$ is a minimizer.
proof of Thm. 18.7.2.

- If \( A \) is a minimizer, then \( A_- \subseteq A \subseteq A_0 \), and \( f(A) = y^*(E) \) is the minimum valuation of \( f \).
- But \( x^* \in P_f \), so \( x^*(A) \leq f(A) \) and \( f(A) = x^*(A_-) \leq x^*(A) \) (or alternatively, just note that \( x^*(A_0 \setminus A) = 0 \)).
- Hence, \( x^*(A) = x^*(A_-) = f(A) \) so that \( A \) is also a tight set for \( x^* \).
- For any \( a \in A \), \( A \) is a tight set containing \( a \), and \( \text{dep}(x^*, a) \) is the minimal tight containing \( a \).
- Hence, for any \( a \in A \), \( \text{dep}(x^*, a) \subseteq A \).
- This means that \( \bigcup_{a \in A} \text{dep}(x^*, a) = A \).
- Since \( A_- \subseteq A \subseteq A_0 \), then \( \exists A_m \subseteq A \setminus A_- \) such that
  \[
  A = \bigcup_{a \in A_-} \text{dep}(x^*, a) \cup \bigcup_{a \in A_m} \text{dep}(x^*, a) = A_- \cup \bigcup_{a \in A_m} \text{dep}(x^*, a)
  \]

Then since \( A \) is a union of tight sets, \( A \) is also a tight set, and we have \( f(A) = x^*(A) \).
- But \( x^*(A \setminus A_-) = 0 \), so \( f(A) = x^*(A) = x^*(A_-) = f(A_-) \) meaning \( A \) is also a minimizer of \( f \).

Therefore, we can generate the entire lattice of minimizers of \( f \) starting from \( A_- \) and \( A_0 \) given access to \( \text{dep}(x^*, e) \).
On a unique minimizer $f$

- Note that if $f(e|A) > 0$, $\forall A \subseteq E$ and $e \in E \setminus A$, then we have $A_− = A_0$ (there is one unique minimizer).
- On the other hand, if $A_− = A_0$, it does not imply $f(e|A) > 0$ for all $A \subseteq E \setminus \{e\}$.
- If $A_− = A_0$ then certainly $f(e|A_0) > 0$ for $e \in E \setminus A_0$ and $-f(e|A_0 \setminus \{e\}) > 0$ for all $e \in A_0$.

Duality: convex minimization of L.E. and min-norm alg.

- Let $f$ be a submodular function with $\tilde{f}$ it’s Lovász extension. Then the following two problems are duals (Bach-2013):

  \[
  \begin{align*}
  \text{minimize} \quad & \tilde{f}(w) + \frac{1}{2}\|w\|_2^2 \\ 
  \text{subject to} \quad & x \in B_f 
  \end{align*}
  \]  
  \hspace{1cm} (18.79) 

  \[
  \begin{align*}
  \text{maximize} \quad & -\|x\|_2^2 \\
  \text{subject to} \quad & x \in B_f 
  \end{align*}
  \]  
  \hspace{1cm} (18.80a) 

  where $B_f = P_f \cap \{x \in \mathbb{R}^V : x(V) = f(V)\}$ is the base polytope of submodular function $f$, and $\|x\|_2^2 = \sum_{e \in V} x(e)^2$ is squared 2-norm.

- Equation (18.79) is related to proximal methods to minimize the Lovász extension (see Parikh&Boyd, “Proximal Algorithms” 2013).

- Equation (18.80b) is solved by the minimum-norm point algorithm (Wolfe-1976, Fujishige-1984, Fujishige-2005, Fujishige-2011) is (as we will see) essentially an active-set procedure for quadratic programming, and uses Edmonds’s greedy algorithm to make it efficient.

- Unknown worst-case running time, although in practice it usually performs quite well (see below).
Convex and affine point hulls, affinely independent

- Given points set \( P = \{p_1, p_2, \ldots, p_k\} \) with \( p_i \in \mathbb{R}^V \), let \( \text{conv} \ P \) be the convex hull of \( P \), i.e.,
  \[
  \text{conv} \ P \triangleq \left\{ \sum_{i=1}^{k} \lambda_i p_i : \sum_{i} \lambda_i = 1, \lambda_i \geq 0, i \in [k] \right\}. \tag{18.81}
  \]

- For a set of points \( Q = \{q_1, q_2, \ldots, q_k\} \), with \( q_i \in \mathbb{R}^V \), we define \( \text{aff} \ Q \) to be the affine hull of \( Q \), i.e.,
  \[
  \text{aff} \ Q \triangleq \left\{ \sum_{i \in 1}^{k} \lambda_i q_i : \sum_{i=1}^{k} \lambda_i = 1 \right\} \supseteq \text{conv} \ Q. \tag{18.82}
  \]

- A set of points \( Q \) is **affinely independent** if no point in \( Q \) belongs to the affine hull of the remaining points.

**H(\( x \)): Orthogonal \( x \)-containing hyperplane**

- Define \( H(\( x \)) \) as the hyperplane that is orthogonal to the line from 0 to \( x \), while also containing \( x \), i.e.
  \[
  H(\( x \)) \triangleq \left\{ y \in \mathbb{R}^V \mid x^\top y = \|x\|^2 \right\}. \tag{18.83}
  \]

- Any set \( \{y \in \mathbb{R}^V | x^\top y = c\} \) is orthogonal to the line from 0 to \( x \). This follows since, for constant \( z \), \( \{y : (y - z)^\top x = 0\} = \{y : y^\top x = z^\top x\} \) is hyperplane orthogonal to \( x \) translated by \( z \). Take \( c = z^\top x \) for result, and \( z = x \), giving \( c = \|x\|^2 \), to contain \( x \).

- Note, \( H(\( x \)) \) is translation of subspace of dimension \( |V| - 1 = n - 1 \) (i.e., \( H(\( x \)) - \{x\} \) is a subspace, \( H(\( x \)) \) is an affine set).
Ex: $H(x)$, polytopes, and supporting hyperplanes

- $H(x) = \{ y \in \mathbb{R}^V | x^T y = \|x\|_2^2 \}$, any $z \in H(x)$ has $x^T z = x^T x$.

- Consider $\text{conv } P$ polytope for points $P = \{p_1, p_2, \ldots \}$, and $\hat{p} \in \text{argmin}_{p \in P} x^T p$. TL: $x^T p < x^T \hat{x}$; TR: $x^T p > x^T \hat{x}$; middle row: $x^T p = x^T \hat{x}$.

- Bottom Row: In Algo, $x$ is chosen so that if $x^T \hat{p} = x^T \hat{x}$ then $H(x)$ separates $P$ from the origin, and $x$ is the min 2-norm point. Notice that $x^T p \geq x^T \hat{x}$ for all $p \in P$.

- Middle/bottom row: $H(x)$ is a supporting hyperplane of $\text{conv } P$ (contained, touching).

Notation

- The line between $x$ and $y$: given two points $x, y \in \mathbb{R}^V$, let $[x, y] \triangleq \{ \lambda x + (1 - \lambda y) : \lambda \in [0, 1] \}$. Hence, $[x, y] = \text{conv } \{x, y\}$.

- Note, if we wish to minimize the 2-norm of a vector $\|x\|_2$, we can equivalently minimize its square $\|x\|_2^2 = \sum_i x_i^2$, and vice versa.
Fujishige-Wolfe Min-Norm Algorithm

- Wolfe-1976 (“Finding the Nearest Point in a Polytope”) developed an algorithm to compute the minimum norm point of a polytope, specified as a set of vertices.
- Fujishige-1984 “Submodular Systems and Related Topics” realized this algorithm can find the min. norm point of \(B_f\).
- Seems to be (among) the fastest general purpose SFM algo.
- Given set of points \(P = \{p_1, \cdots, p_m\}\) where \(p_i \in \mathbb{R}^n\): find the minimum norm point in convex hull of \(P\):
  \[
  \min_{x \in \text{conv} P} \|x\|_2 
  \]  
  (18.84)
- Wolfe’s algorithm is guaranteed terminating, and explicitly uses a representation of \(x\) as a convex combination of points in \(P\).
- Algorithm maintains a set of points \(Q \subseteq P\), which is always assuredly affinely independent.

When \(Q\) are affinely independent, minimum norm point in the affine hull of \(Q\) can easily be found, as a closed form solution for \(\min_{x \in \text{aff} Q} \|x\|_2\) is available (see below).

Algorithm repeatedly produces min. norm point \(x^*\) for selected set \(Q\).

If we find \(w_i \geq 0, i = 1, \cdots, m\) for the minimum norm point, then \(x^*\) also belongs to \(\text{conv} Q\) and also a minimum norm point over \(\text{conv} Q\).

If \(Q \subseteq P\) is suitably chosen, \(x^*\) may even be the minimum norm point over \(\text{conv} P\) solving the original problem.

One of the most expensive parts of Wolfe’s algorithm is solving linear optimization problem over the polytope, doable by examining all the extreme points in the polytope.

If number of extreme points is exponential, hard to do in general.

Number of extreme points of submodular base polytope is exponentially large, but linear optimization over the base polytope \(B_f\) doable \(O(n \log n)\) time via Edmonds’s greedy algorithm.
Pseudocode of Fujishige-Wolfe Min-Norm (MN) algorithm

**Input**: \( P = \{ p_1, \ldots, p_m \}, p_i \in \mathbb{R}^n, i = 1, \ldots, m \).

**Output**: \( x^* \): the minimum-norm-point in \( \text{conv} \ P \).

1. \( x^* \leftarrow p_i^* \) where \( p_i^* \in \arg\min_{p \in P} \| p \|_2 \) /* or choose it arbitrarily */;
2. \( Q \leftarrow \{ x^* \} \);
3. while 1 do /* major loop */
   4. if \( x^* = 0 \) or \( H(x^*) \) separates \( P \) from origin then
      return : \( x^* \)
   5. else
      6. Choose \( \hat{x} \in P \) on the near (closer to 0) side of \( H(x^*) \);
      7. \( Q = Q \cup \{ \hat{x} \} \);
      8. while 1 do /* minor loop */
         9. \( x_0 \leftarrow \arg\min_{x \in \text{aff} Q} \| x \|_2 \);
         10. if \( x_0 \in \text{conv} Q \) then
             11. \( x^* \leftarrow x_0 \);
             12. break;
         13. else
             14. \( y \leftarrow \arg\min_{x \in \text{conv} Q \cap [x^*, x_0]} \| x - x_0 \|_2 \);
             15. Delete from \( Q \) points not on the face of \( \text{conv} Q \) where \( y \) lies;
             16. \( x^* \leftarrow y \);

Fujishige-Wolfe Min-Norm algorithm: Geometric Example

- It is advised that for the next set of slides, you have a print out of the previous MN algorithm available on display/paper somewhere.
- Algorithm maintains an invariant, namely that:
  \[ x^* \in \text{conv} Q \subseteq \text{conv} P, \]  
  (18.85)

must hold at every possible assignment of \( x^* \) (Lines 1, 11, and 16):
  1. True after Line 1 since \( Q = \{ x^* \} \),
  2. True after Line 11 since \( x_0 \in \text{conv} Q \),
  3. and true after Line 16 since \( y \in \text{conv} Q \) even after deleting points.
- Note also for any \( x^* \in \text{conv} Q \subseteq \text{conv} P \), we have
  \[ \min_{x \in \text{aff} Q} \| x \|_2 \leq \min_{x \in \text{conv} Q} \| x \|_2 \leq \| x^* \|_2 \]  
  (18.86)

- Note, the input, \( P \), consists of \( m \) points. In the case of the base polytope, \( P = B_f \) could be exponential in \( n = |V| \).
- There are six places that might be seemingly tricky or expensive: Line 4, Line 6, Line 9, Line 10, Line 14, and Line 15.
- We will consider each in turn, but first we do a geometric example.
Polytope, and circles concentric at 0.

The initial polytope consisting of the convex hull of three points $p_1, p_2, p_3$, and the origin 0.
$p_1$ is the extreme point closest to 0 and so we choose it first, although we can choose any arbitrary extreme point as the initial point. We set $x^* \leftarrow p_1$ in Line 1, and $Q \leftarrow \{p_1\}$ in Line 2. $H(x^*) = H(p_1)$ (green dashed line) is not a supporting hyperplane of $\text{conv}(P)$ in Line 4, so we move on to the else condition in Line 5.

We need to add some extreme point $\hat{x}$ on the “near” side of $H(p_1)$ in Line 6, we choose $\hat{x} = p_2$. In Line 7, we set $Q \leftarrow Q \cup \{p_2\}$, so $Q = \{p_1, p_2\}$. 
$x_0 = R$ is the min-norm point in $\text{aff}\{p_1, p_2\}$ computed in Line 9. Also, with $Q = \{p_1, p_2\}$, since $R \in \text{conv} Q$, we set $x^* \leftarrow x_0 = R$ in Line 11, not violating the invariant $x^* \in \text{conv} Q$. Note, after Line 11, we still have $x^* \in \text{conv} P$ and $\|x^*\|_2 = \|x^*_\text{new}\|_2 < \|x^*_\text{old}\|_2$ strictly.
R = x_0 = x^\ast. We consider next H(R) = H(x^\ast) in Line 4. H(x^\ast) is not a supporting hyperplane of conv P. So we choose p_3 on the “near” side of H(x^\ast) in Line 6. Add Q ← Q \cup \{p_3\} in Line 7. Now Q = P = \{p_1, p_2, p_3\}. The origin x_0 = 0 is the min-norm point in aff Q (Line 9), and it is not in the interior of conv Q (condition in Line 10 is false).
Fujishige-Wolfe Min-Norm algorithm: Geometric Example

\[ Q = P = \{p_1, p_2, p_3\} \]. Line 14: \( S = y = \arg\min_{x \in \text{conv } Q \cap [x^*, x_0]} \| x - x_0 \|_2 \) where \( x_0 \) is 0 and \( x^* \) is \( R \) here. Thus, \( y \) lies on the boundary of \( \text{conv } Q \). Note, \( \| y \|_2 < \| x^* \|_2 \) since \( x^* \in \text{conv } Q \), \( \| x_0 \|_2 < \| x^* \|_2 \). Line 15: Delete \( p_1 \) from \( Q \) since not on face where \( y = S \) lies. \( Q = \{p_2, p_3\} \) after Line 15. We still have \( y = S \in \text{conv } Q \) for the updated \( Q \). Line 16: \( x^* \leftarrow y \), retain invariant \( x^* \in \text{conv } Q \), and again have \( \| x^* \|_2 = \| x^*_{\text{new}} \|_2 < \| x^*_{\text{old}} \|_2 \) strictly.

\[ Q = \{p_2, p_3\}, \text{ and so } x_0 = T \text{ computed in Line 9 is the min-norm point in } \text{aff } Q \]. We also have \( x_0 \in \text{conv } Q \) in Line 10 so we assign \( x^* \leftarrow x_0 \) in Line 11 and break.
Fujishige-Wolfe Min-Norm algorithm: Geometric Example

$H(T)$ separates $P$ from the origin in Line 4, and therefore is a supporting hyperplane, and therefore $x^*$ is the min-norm point in conv $P$, so we return with $x^*$.

**Condition for Min-Norm Point**

**Theorem 18.8.1**

$P = \{p_1, p_2, \ldots, p_m\}$, $x^* \in \text{conv} P$ is the min. norm point in conv $P$ iff

$$p_i^\top x^* \geq \| x^* \|_2^2 \quad \forall i = 1, \cdots, m.$$  \hspace{1cm} (18.87)

**Proof.**

- Assume $x^*$ is the min-norm point, let $y \in \text{conv} P$, and $0 \leq \theta \leq 1$.
- Then $z \triangleq x^* + \theta(y - x^*) = (1 - \theta) x^* + \theta y \in \text{conv} P$, and
  $$\| z \|_2^2 = \| x^* + \theta(y - x^*) \|_2^2 = \| x^* \|_2^2 + 2\theta(x^\top y - x^\top x^*) + \theta^2 \| y - x^* \|_2^2 \quad \text{(18.88)}$$
  $$\| z \|_2^2 > \| x^* \|_2^2 \quad \text{for arbitrary } z \in \text{conv} P \Rightarrow \text{Equation (18.87).}$$

- It is possible for $\| z \|_2^2 < \| x^* \|_2^2$ for small $\theta$, unless $x^\top y \geq x^\top x^*$ for all $y \in \text{conv} P$.
- Conversely, given Eq (18.87), and given that $y = \sum_i \lambda_i p_i \in \text{conv} P$,
  $$y^\top x^* = \sum_i \lambda_i p_i^\top x^* \geq \sum_i \lambda_i x^\top x^* = x^\top x^* \hspace{1cm} (18.90)$$
  implying that $\| z \|_2^2 > \| x^* \|_2^2$ in Equation 18.89 for arbitrary $z \in \text{conv} P$. 

---

**Prof. Jeff Bilmes**
EE563/Spring 2018/Submodularity - Lecture 18 - May 30th, 2018
F68/85 (pg.77/94)
The set $Q$ is always affinely independent

**Lemma 18.8.2**

The set $Q$ in the MN Algorithm is always affinely independent.

**Proof.**

- $Q$ is of course affinely independent when there is at most one point in it (e.g., after Line 2).
- After the initialization, it changes only by deletion of points, or adding a single point. Deletion does not change the independence.
- Before adding $\hat{x}$ at Line 7, we know $x^*$ is the minimum norm point in aff $Q$ (since we break only at Line 12).
- Therefore, $x^*$ is normal to aff $Q$, which implies aff $Q \subseteq H(x^*)$.
- Since $\hat{x} \notin H(x^*)$ chosen at Line 6, we have $\hat{x} \notin$ aff $Q$.
- ∴ update $Q \cup \{\hat{x}\}$ at Line 7 is affinely independent as long as $Q$ is.

Thus, by Lemma 18.8.2, we have for any $x \in$ aff $Q$ such that $x = \sum_i w_i q_i$ with $\sum_i w_i = 1$, the weights $w_i$ are uniquely determined.

The set $Q$ is never too large

**Lemma 18.8.3**

The set $Q$ in the MN Algorithm has size never more than $n + 1$.

**Proof.**

This is immediate, since $Q$ is always affinely independent, and in $\mathbb{R}^V$, an affinely independent set can have at most $n + 1$ entries, with $|V| = n$. □
Minimum Norm in an affine set

- Line 9 of the algorithm requires $x_0 \leftarrow \min_{x \in \text{aff } Q} \|x\|_2$.
- When $Q$ is affinely independent, this is relatively easy.
- Let $Q$ represent $n \times k$ matrix with points as columns $q \in Q$. The following is solvable with matrix inversion/linear solver, where $x = Qw$:

  minimize $\|x\|_2^2 = w^TQ^TQw$ \hspace{1cm} (18.91)

  subject to $1^T w = 1$ \hspace{1cm} (18.92)

- Form Lagrangian $w^TQ^TQw + 2\lambda(1^Tw - 1)$, and differentiating w.r.t. $\lambda$ and $w$, and setting to zero, we get:

  $1^T w = 1$ \hspace{1cm} (18.93)

  $Q^TQw + \lambda 1 = 0$ \hspace{1cm} (18.94)

- $k + 1$ variables and $k$ unknowns, solvable with linear solver with matrices

  $\begin{bmatrix} 0 & 1^T \\ 1 & Q^TQ \end{bmatrix} \begin{bmatrix} \lambda \\ w \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ \hspace{1cm} (18.95)

- Thanks to $Q$ being affine, matrix on l.h.s. is invertable.

Minimum Norm in an affine set

- Note, this also solves Line 10, since feasibility requires $\sum_i w_i = 1$, we need only check $w \geq 0$ to ensure $x_0 = \sum_i w_i q_i \in \text{conv } Q$.

- In fact, a feature of the algorithm (in Wolfe’s 1976 paper) is that we keep the convex coefficients $\{w_i\}_i$ where $x^* = \sum_i w_ip_i$ of $x^*$ and from this vector. We also keep $v$ such that $x_0 = \sum_i v_i q_i$ for points $q_i \in Q$, from Line 9.

- Given $w$ and $v$, we can also easily solve Lines 14 and 15 (see “Step 3” on page 133 of Wolfe-1976, which also defines numerical tolerances).

- We have yet to see how to efficiently solve Lines 4 and 6, however.
The MN Algorithm finds the minimum norm point in $\text{conv} \ P$ after a finite number of iterations of the major loop.

Proof.

- In minor loop, we always have $x^* \in \text{conv} \ Q$, since whenever $Q$ is modified, $x^*$ is updated as well (Line 16) such that the updated $x^*$ remains in new $\text{conv} \ Q$.

- Hence, every time $x^*$ is updated (in minor loop), its norm never increases, i.e., before Line 11, $\|x_0\|_2 \leq \|x^*\|_2$ since $x^* \in \text{aff} \ Q$ and $x_0 = \min_{x \in \text{aff} \ Q} \|x\|_2$. Similarly, before Line 16, $\|y\|_2 \leq \|x^*\|_2$, since invariant $x^* \in \text{conv} \ Q$ but while $x_0 \in \text{aff} \ Q$, we have $x_0 \notin \text{conv} \ Q$, and $\|x_0\|_2 < \|x^*\|_2$.

Moreover, there can be no more iterations within a minor loop than the dimension of $\text{conv} \ Q$ for the initial $Q$ given to the minor loop initially at Line 8 (dimension of $\text{conv} \ Q$ is $|Q| - 1$ since $Q$ is affinely independent).

- Each iteration of the minor loop removes at least one point from $Q$ in Line 15.

- When $Q$ reduces to a singleton, the minor loop always terminates.

- Thus, the minor loop terminates in finite number of iterations, at most dimension of $Q$.

- In fact, total number of iterations of minor loop in entire algorithm is at most number of points in $P$ since we never add back in points to $Q$ that have been removed.
MN Algorithm finds the MN point in finite time.

...proof of Theorem 18.8.4 continued.

- Each time $Q$ is augmented with $\hat{x}$ at Line 7, followed by updating $x^*$ with $x_0$ at Line 11, (i.e., when the minor loop returns with only one iteration), $\|x^*\|_2$ strictly decreases from what it was before.

- To see this, consider $x^* + \theta(\hat{x} - x^*)$ where $0 \leq \theta \leq 1$. Since both $\hat{x}, x^* \in \text{conv } Q$, we have $x^* + \theta(\hat{x} - x^*) \in \text{conv } Q$.

- Therefore, we have $\|x^* + \theta(\hat{x} - x^*)\|_2 \geq \|x_0\|_2$, which implies

\[
\|x^* + \theta(\hat{x} - x^*)\|_2^2 = \|x^*\|_2^2 + 2\theta \left( (x^*)^T \hat{x} - \|x^*\|_2 \right) + \theta^2 \|\hat{x} - x^*\|_2^2 \\
\geq \|x_0\|_2^2
\]  

(18.96)

and from Line 6, $\hat{x}$ is on the same side of $H(x^*)$ as the origin, i.e. $(x^*)^T \hat{x} < \|x^*\|_2^2$, so middle term of r.h.s. of equality is negative.

...
Line: 6: Finding $\hat{x} \in P$ on the near side of $H(x^*)$

- The “near” side means the side that contains the origin.
- Ideally, find $\hat{x}$ such that the reduction of $\|x^*\|_2$ is maximized to reduce number of major iterations.
- From Eqn. 18.96, reduction on norm is lower-bounded:

$$\Delta = \|x^*\|_2^2 - \|x_0\|_2^2 \geq 2\theta \left( \|x^*\|_2^2 - (x^*)^\top \hat{x} \right) - \theta^2 \|\hat{x} - x^*\|_2^2 \equiv \Delta$$  \hspace{0.5cm} (18.98)

- When $0 \leq \theta < \frac{2(\|x^*\|_2^2 - (x^*)^\top \hat{x})}{\|\hat{x} - x^*\|_2^2}$, we can get the maximal value of the lower bound, over $\theta$, as follows:

$$\max_{0 \leq \theta < \frac{2(\|x^*\|_2^2 - (x^*)^\top \hat{x})}{\|\hat{x} - x^*\|_2^2}} \Delta = \left( \frac{\|x^*\|_2^2 - (x^*)^\top \hat{x}}{\|\hat{x} - x^*\|_2^2} \right)^2$$ \hspace{0.5cm} (18.99)

To maximize lower bound of norm reduction at each major iteration, want to find an $\hat{x}$ such that the above lower bound (Equation 18.99) is maximized.

That is, we want to find

$$\hat{x} \in \arg\max_{x \in P} \left( \frac{\|x^*\|_2^2 - (x^*)^\top x}{\|x - x^*\|_2} \right)^2$$ \hspace{0.5cm} (18.100)

to ensure that a large norm reduction is assured.

This problem, however, is at least as hard as the MN problem itself as we have a quadratic term in the denominator.
Line: 6: Finding $\hat{x} \in P$ on the near side of $H(x^*)$

- As a surrogate, we maximize numerator in Eqn. 18.100, i.e., find

$$\hat{x} \in \arg\max_{x \in P} \|x^*\|^2 - (x^*)^T x = \arg\min_{x \in P} (x^*)^T x,$$

(18.101)

- Intuitively, by solving the above, we find $\hat{x}$ such that it has the largest “distance” to the hyperplane $H(x^*)$, and this is exactly the strategy used in the Wolfe-1976 algorithm.

- Also, solution $\hat{x}$ in Line 6 can be used to determine if hyperplane $H(x^*)$ separates $\text{conv} P$ from the origin (Line 4): if the point in $P$ having greatest distance to $H(x^*)$ is not on the side where origin lies, then $H(x^*)$ separates $\text{conv} P$ from the origin.

- Mathematically and theoretically, we terminate the algorithm if

$$((x^*)^T \hat{x} \geq \|x^*\|^2),$$

(18.102)

where $\hat{x}$ is the solution of Eq. 18.101.

In practice, the above optimality test might never hold numerically. Hence, as suggested by Wolfe, we introduce a tolerance parameter $\epsilon > 0$, and terminates the algorithm if

$$(x^*)^T \hat{x} > \|x^*\|^2 - \epsilon \max_{x \in Q} \|x\|^2$$

(18.103)

- When $\text{conv} P$ is a submodular base polytope (i.e., $\text{conv} P = B_f$ for a submodular function $f$), then the problem in Eqn 18.101 can be solved efficiently by Edmonds’s greedy algorithm (even though there may be an exponential number of extreme points).

- Edmond’s greedy algorithm, therefore, solves both Line 4 and Line 6 simultaneously.

- Hence, Edmonds’s discovery is one of the main reasons that the MN algorithm is applicable to submodular function minimization.
The currently fastest strongly polynomial combinatorial algorithm for SFM achieves a running time of $O(n^5T + n^6)$ (Orlin’09) where $T$ is the time for function evaluation, far from practical for large problem instances.

Fujishige & Isotani report that MN algorithm is fast in practice, but they use only a limited set of submodular functions.

- Complexity of MN Algorithm is still an unsolved problem.
- Obvious facts:
  - each major iteration requires $O(n)$ function oracle calls
  - complexity of each major iteration could be at least $O(n^3)$ due to the affine projection step (solving a linear system).
  - Therefore, the complexity of each major iteration is $O(n^3 + n^{1+p})$

where each function oracle call requires $O(n^p)$ time.

Since the number of major iterations required is unknown, the complexity of MN is also unknown.
MN Algorithm Empirical Complexity

- A lower bound complexity of the min-norm has not been established.
- In 2014, Chakrabarty, Jain, and Kothari in their NIPS 2014 paper “Provable Submodular Minimization using Wolfe’s Algorithm” showed a pseudo-polynomial time bound of $O(n^7 g_f^2)$ where $n = |V|$ is the ground set, and $g_f$ is the maximum gain of a particular function $f$.
- This is pseudo-polynomial since it depends on the function values.
- There currently is no known polynomial time complexity analysis for this algorithm.