Submodular Functions, Optimization, and Applications to Machine Learning
— Spring Quarter, Lecture 17 —

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 23st, 2018

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

- \(f(A) + 2f(C) + f(B)\)
- \(f(A) + f(C) + f(B)\)
- \(f(A)\)
Announcements, Assignments, and Reminders

- Next homework will be posted tonight.
- Rest of the quarter. One more longish homework.
- Take home final exam (like a long homework).
- As always, if you have any questions about anything, please ask then 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).
Class Road Map - EE563

L1(3/26): Motivation, Applications, & Basic Definitions,
L3(4/2): Info theory exs, more apps, definitions, graph/combinatorial examples
L4(4/4): Graph and Combinatorial Examples, Matrix Rank, Examples and Properties, visualizations
L5(4/9): More Examples/Properties/Other SubmodularDefs., Independence,
L6(4/11): Matroids, Matroid Examples, Matroid Rank, Partition/Laminar Matroids
L7(4/16): Laminar Matroids, System of Distinct Reps, Transversals, Transversal Matroid, Matroid Representation, Dual Matroids
L8(4/18): Dual Matroids, Other Matroid Properties, Combinatorial Geometries, Matroids and Greedy.
L9(4/23): Polyhedra, Matroid Polytopes, Matroids → Polymatroids
L10(4/29): Matroids → Polymatroids, Polymatroids, Polymatroids and Greedy,
L11(4/30): Polymatroids, Polymatroids and Greedy
L12(5/2): Polymatroids and Greedy, Extreme Points, Cardinality Constrained Maximization
L13(5/7): Constrained Submodular Maximization
L14(5/9): Submodular Max w. Other Constraints, Cont. Extensions, Lovasz Extension
L16(5/16): More Lovasz extension, Choquet, defs/props, examples, multilinear extension
L17(5/21): Finish L.E., Multilinear Extension, Submodular Max/polyhedral approaches, Most Violated inequality, Still More on Matroids, Closure/Sat
L18(5/23):
L–(5/28): Memorial Day (holiday)
L19(5/30):

Last day of instruction, June 1st. Finals Week: June 2-8, 2018.
One slide review of concave relaxation

convex closure $\tilde{f}(x) = \min_{p \in \triangle^n(x)} ES \sim_p [f(S)]$, where $\triangle^n(x) = \{ p \in \mathbb{R}^{2^n} : \sum_{S \subseteq V} p_S = 1, p_S \geq 0 \forall S \subseteq V, \& \sum_{S \subseteq V} p_S 1_S = x \}$

“Edmonds” extension $\bar{f}(w) = \max(wx : x \in B_f)$

Lovász extension $f_{LE}(w) = \sum_{i=1}^{m} \lambda_i f(E_i)$, with $\lambda_i$ such that $w = \sum_{i=1}^{m} \lambda_i 1_{E_i}$

$\tilde{f}(w) = \max_{\sigma \in \Pi_{[m]}} w^\top c^\sigma$, $\Pi_{[m]}$ set of $m!$ permutations of $[m]$, $\sigma \in \Pi_{[m]}$ a permutation, $c^\sigma$ vector with $c^\sigma_i = f(E_{\sigma i}) - f(E_{\sigma i-1})$, $E_{\sigma i} = \{e_{\sigma 1}, e_{\sigma 2}, \ldots, e_{\sigma i}\}$.

Choquet integral $C_f(w) = \sum_{i=1}^{m} (w_{e_i} - w_{e_{i+1}}) f(E_i)$

$\bar{f}(w) = \int_{-\infty}^{+\infty} \hat{f}(\alpha) d\alpha, \hat{f}(\alpha) = \begin{cases} f(\{w \geq \alpha\}) & \text{if } \alpha \geq 0 \\ f(\{w \geq \alpha\}) - f(E) & \text{if } \alpha < 0 \end{cases}$

All the same when $f$ is submodular.
Lovász extension properties

- Using the above, have the following (some of which we’ve seen):

### Theorem 17.2.2

Let \( f, g : 2^E \to \mathbb{R} \) be normalized \((f(\emptyset) = g(\emptyset) = 0)\). Then

1. **Superposition of LE operator**: Given \( f \) and \( g \) with Lovász extensions \( \tilde{f} \) and \( \tilde{g} \) then \( \tilde{f} + \tilde{g} \) is the Lovász extension of \( f + g \) and \( \lambda \tilde{f} \) is the Lovász extension of \( \lambda f \) for \( \lambda \in \mathbb{R} \).

2. If \( w \in \mathbb{R}_+^E \) then \( \tilde{f}(w) = \int_0^\infty f(\{w \geq \alpha\})d\alpha \).

3. For \( w \in \mathbb{R}^E \), and \( \alpha \in \mathbb{R} \), we have \( \tilde{f}(w + \alpha 1_E) = \tilde{f}(w) + \alpha f(E) \).

4. **Positive homogeneity**: I.e., \( \tilde{f}(\alpha w) = \alpha \tilde{f}(w) \) for \( \alpha \geq 0 \).

5. For all \( A \subseteq E, \tilde{f}(1_A) = f(A) \).

6. \( f \) symmetric as in \( f(A) = f(E \setminus A), \forall A \), then \( \tilde{f}(w) = \tilde{f}(-w) \) \((\tilde{f} \text{ is even})\).

7. Given partition \( E^1 \cup E^2 \cup \cdots \cup E^k \) of \( E \) and \( w = \sum_{i=1}^k \gamma_i 1_{E_i} \), with \( \gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_k \), and with \( E^{1:i} = E^1 \cup E^2 \cup \cdots \cup E^i \), then \( \tilde{f}(w) = \sum_{i=1}^k \gamma_i f(E^i | E^{1:i-1}) = \sum_{i=1}^{k-1} f(E^{1:i})(\gamma_i - \gamma_{i+1}) + f(E)\gamma_k \).
Example: \( m = 3, \ E = \{1, 2, 3\} \)

- In order to visualize in 3D, we make a few simplifications.
Example: \( m = 3, \ E = \{1, 2, 3\} \)

- In order to visualize in 3D, we make a few simplifications.
- Consider any submodular \( f' \) and \( x \in B_{f'} \). Then \( f(A) = f'(A) - x(A) \) is submodular.
Example: $m = 3, E = \{1, 2, 3\}$

- In order to visualize in 3D, we make a few simplifications.
- Consider any submodular $f'$ and $x \in B_{f'}$. Then $f(A) = f'(A) - x(A)$ is submodular, and moreover $f(E) = f'(E) - x(E) = 0$. 
Example: \( m = 3, \ E = \{1, 2, 3\} \)

- In order to visualize in 3D, we make a few simplifications.
- Consider any submodular \( f' \) and \( x \in B_{f'} \). Then \( f(A) = f'(A) - x(A) \) is submodular, and moreover \( f(E) = f'(E) - x(E) = 0 \).
- Hence, from \( \tilde{f}(w + \alpha 1_E) = \tilde{f}(w) + \alpha f(E) \), we have that \( \tilde{f}(w + \alpha 1_E) = \tilde{f}(w) \) when \( f(E) = 0 \).
Example: $m = 3, \ E = \{1, 2, 3\}$

- In order to visualize in 3D, we make a few simplifications.
- Consider any submodular $f'$ and $x \in B_{f'}$. Then $f(A) = f'(A) - x(A)$ is submodular, and moreover $f(E) = f'(E) - x(E) = 0$.
- Hence, from $\tilde{f}(w + \alpha 1_E) = \tilde{f}(w) + \alpha f(E)$, we have that $\tilde{f}(w + \alpha 1_E) = \tilde{f}(w)$ when $f(E) = 0$.
- Thus, we can look “down” on the contour plot of the Lovász extension, $\{w : \tilde{f}(w) = 1\}$, from a vantage point right on the line $\{x : x = \alpha 1_E, \alpha > 0\}$ since moving in direction $1_E$ changes nothing.
Example: \( m = 3, \ E = \{1, 2, 3\} \)

- In order to visualize in 3D, we make a few simplifications.
- Consider any submodular \( f' \) and \( x \in B_{f'} \). Then \( f(A) = f'(A) - x(A) \) is submodular, and moreover \( f(E) = f'(E) - x(E) = 0 \).
- Hence, from \( \tilde{f}(w + \alpha \mathbf{1}_E) = \tilde{f}(w) + \alpha f(E) \), we have that \( \tilde{f}(w + \alpha \mathbf{1}_E) = \tilde{f}(w) \) when \( f(E) = 0 \).
- Thus, we can look “down” on the contour plot of the Lovász extension, \( \{w : \tilde{f}(w) = 1\} \), from a vantage point right on the line \( \{x : x = \alpha \mathbf{1}_E, \alpha > 0\} \) since moving in direction \( \mathbf{1}_E \) changes nothing.
- I.e., consider 2D plane perpendicular to the line \( \{x : \exists \alpha, x = \alpha \mathbf{1}_E\} \) at any point along that line, then Lovász extension is surface plot with coordinates on that plane (or alternatively we can view contours).
Example: \( m = 3, \ E = \{1, 2, 3\} \)

- Example 1 (from Bach-2011): \( f(A) = \mathbf{1}_{|A| \in \{1, 2\}} \)
  \[
  = \min \{|A|, 1\} + \min \{|E \setminus A|, 1\} - 1 
  \]
  is submodular, and \( \tilde{f}(w) = \max_{k \in \{1, 2, 3\}} w_k - \min_{k \in \{1, 2, 3\}} w_k. \)
Example: $m = 3$, $E = \{1, 2, 3\}$

- Example 1 (from Bach-2011): $f(A) = 1_{|A| \in \{1, 2\}}$
  
  \[
  f(A) = \min \{|A|, 1\} + \min \{|E \setminus A|, 1\} - 1 \text{ is submodular, and} 
  \tilde{f}(w) = \max_{k \in \{1, 2, 3\}} w_k - \min_{k \in \{1, 2, 3\}} w_k.
  \]

\[
\begin{align*}
(1,0,1)/F(\{1,3\}) & \quad (0,0,1)/F(\{3\}) \\
(1,0,0)/F(\{1\}) & \quad (0,1,0)/F(\{2\}) \\
(1,0,1)/F(\{1,3\}) & \quad (0,1,1)/F(\{2,3\})
\end{align*}
\]
Example: $m = 3$, $E = \{1, 2, 3\}$

Example 2 (from Bach-2011): $f(A) = |1_{1 \in A} - 1_{2 \in A}| + |1_{2 \in A} - 1_{3 \in A}|$

This gives a "total variation" function for the Lovász extension, with $\tilde{f}(w) = |w_1 - w_2| + |w_2 - w_3|$.

When used as a prior, prefers piecewise-constant signals (e.g., $P_i | w_i | w_i + 1 |$).
Example: $m = 3, \ E = \{1, 2, 3\}$

- Example 2 (from Bach-2011): $f(A) = |1_{1 \in A} - 1_{2 \in A}| + |1_{2 \in A} - 1_{3 \in A}|$

This gives a “total variation” function for the Lovász extension, with $\tilde{f}(w) = |w_1 - w_2| + |w_2 - w_3|$. 
Example: \( m = 3, \ E = \{1, 2, 3\} \)

- Example 2 (from Bach-2011): \( f(A) = \left| \mathbf{1}_{1 \in A} - \mathbf{1}_{2 \in A} \right| + \left| \mathbf{1}_{2 \in A} - \mathbf{1}_{3 \in A} \right| \)
  
  This gives a “total variation” function for the Lovász extension, with 
  \( \hat{f}(w) = |w_1 - w_2| + |w_2 - w_3| \).

- When used as a prior, prefers piecewise-constant signals (e.g., \( \sum_i |w_i - w_{i+1}| \)).
Total Variation Example

From “Nonlinear total variation based noise removal algorithms” Rudin, Osher, and Fatemi, 1992. Top left original, bottom right total variation.
Example: Lovász extension of concave over modular

Let $m : E \rightarrow \mathbb{R}_+$ be a modular function and define $f(A) = g(m(A))$ where $g$ is concave. Then $f$ is submodular.
Example: Lovász extension of concave over modular

- Let $m : E \rightarrow \mathbb{R}_+$ be a modular function and define $f(A) = g(m(A))$ where $g$ is concave. Then $f$ is submodular.
- Let $M_j = \sum_{i=1}^{j} m(e_i)$
Example: Lovász extension of concave over modular

- Let \( m : E \to \mathbb{R}_+ \) be a modular function and define \( f(A) = g(m(A)) \) where \( g \) is concave. Then \( f \) is submodular.
- Let \( M_j = \sum_{i=1}^{j} m(e_i) \)
- \( \tilde{f}(w) \) is given as

\[
\tilde{f}(w) = \sum_{i=1}^{m} w(e_i)(g(M_i) - g(M_{i-1}))
\]  

(17.1)
Example: Lovász extension of concave over modular

Let $m : E \to \mathbb{R}_+$ be a modular function and define $f(A) = g(m(A))$ where $g$ is concave. Then $f$ is submodular.

Let $M_j = \sum_{i=1}^{j} m(e_i)$

$\tilde{f}(w)$ is given as

$$
\tilde{f}(w) = \sum_{i=1}^{m} w(e_i)(g(M_i) - g(M_{i-1})) \quad (17.1)
$$

And if $m(A) = |A|$, we get

$$
\tilde{f}(w) = \sum_{i=1}^{m} w(e_i)(g(i) - g(i - 1)) \quad (17.2)
$$
Example: Lovász extension and cut functions

- Cut Function: Given a non-negative weighted graph \( G = (V, E, m) \) where \( m : E \rightarrow \mathbb{R}_+ \) is a modular function over the edges, we know from Lecture 2 that \( f : 2^V \rightarrow \mathbb{R}_+ \) with \( f(X) = m(\Gamma(X)) \) where
  \[
  \Gamma(X) = \{(u, v) | (u, v) \in E, u \in X, v \in V \setminus X\}
  \]
  is non-monotone submodular.
Example: Lovász extension and cut functions

- Cut Function: Given a non-negative weighted graph $G = (V, E, m)$ where $m : E \to \mathbb{R}_+$ is a modular function over the edges, we know from Lecture 2 that $f : 2^V \to \mathbb{R}_+$ with $f(X) = m(\Gamma(X))$ where $\Gamma(X) = \{(u, v) | (u, v) \in E, u \in X, v \in V \setminus X\}$ is non-monotone submodular.

- Simple way to write it, with $m_{ij} = m((i, j))$:

$$f(X) = \sum_{i \in X, j \in V \setminus X} m_{ij}$$  \hspace{1cm} (17.3)
Example: Lovász extension and cut functions

- **Cut Function**: Given a non-negative weighted graph $G = (V, E, m)$ where $m : E \rightarrow \mathbb{R}_+$ is a modular function over the edges, we know from Lecture 2 that $f : 2^V \rightarrow \mathbb{R}_+$ with $f(X) = m(\Gamma(X))$ where $\Gamma(X) = \{(u, v) | (u, v) \in E, u \in X, v \in V \setminus X\}$ is non-monotone submodular.

- Simple way to write it, with $m_{ij} = m((i, j))$:

  $$f(X) = \sum_{i \in X, j \in V \setminus X} m_{ij} \quad (17.3)$$

- **Exercise**: show that Lovász extension of graph cut may be written as:

  $$\tilde{f}(w) = \sum_{i, j \in V} m_{ij} \max \{(w_i - w_j), 0\} \quad (17.4)$$

  where elements are ordered as usual, $w_1 \geq w_2 \geq \cdots \geq w_n$. 
Example: Lovász extension and cut functions

- **Cut Function**: Given a non-negative weighted graph $G = (V, E, m)$ where $m: E \rightarrow \mathbb{R}_+$ is a modular function over the edges, we know from Lecture 2 that $f: 2^V \rightarrow \mathbb{R}_+$ with $f(X) = m(\Gamma(X))$ where $\Gamma(X) = \{(u, v) | (u, v) \in E, u \in X, v \in V \setminus X\}$ is non-monotone submodular.

- **Simple way to write it**, with $m_{ij} = m((i, j))$:

$$f(X) = \sum_{i \in X, j \in V \setminus X} m_{ij} \quad (17.3)$$

- **Exercise**: show that Lovász extension of graph cut may be written as:

$$\tilde{f}(w) = \sum_{i, j \in V} m_{ij} \max \{(w_i - w_j), 0\} \quad (17.4)$$

where elements are ordered as usual, $w_1 \geq w_2 \geq \cdots \geq w_n$.

- **This is also a form of “total variation”**
Some additional submodular functions and their Lovász extensions, where \( w(e_1) \geq w(e_2) \geq \cdots \geq w(e_m) \geq 0 \). Let \( W_k \triangleq \sum_{i=1}^{k} w(e_i) \).

<table>
<thead>
<tr>
<th>( f(A) )</th>
<th>( \tilde{f}(w) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(</td>
<td>A</td>
</tr>
<tr>
<td>( \min(</td>
<td>A</td>
</tr>
<tr>
<td>( \min(</td>
<td>A</td>
</tr>
<tr>
<td>( \min(</td>
<td>A</td>
</tr>
<tr>
<td>( \min(</td>
<td>A</td>
</tr>
<tr>
<td>( \min(</td>
<td>A</td>
</tr>
</tbody>
</table>

(thanks to K. Narayanan).
Supervised And Unsupervised Machine Learning

- Given training data $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^m$ with $(x_i, y_i) \in \mathbb{R}^n \times \mathbb{R}$, perform the following risk minimization problem:

$$
\min_{w \in \mathbb{R}^n} \frac{1}{m} \sum_{i=1}^m \ell(y_i, w^T x_i) + \lambda \Omega(w),
$$

where $\ell(\cdot)$ is a loss function (e.g., squared error) and $\Omega(w)$ is a norm.

- When data has multiple responses $(x_i, y_i) \in \mathbb{R}^n \times \mathbb{R}^k$, learning becomes:

$$
\min_{w^1, \ldots, w^k \in \mathbb{R}^n} \sum_{j=1}^k \frac{1}{m} \sum_{i=1}^m \ell(y_i^k, (w^k)^T x_i) + \lambda \Omega(w^k),
$$

- When data has multiple responses only that are observed, $(y_i) \in \mathbb{R}^k$, we get dictionary learning (Krause & Guestrin, Das & Kempe):

$$
\min_{x_1, \ldots, x_m} \min_{w^1, \ldots, w^k \in \mathbb{R}^n} \sum_{j=1}^k \frac{1}{m} \sum_{i=1}^m \ell(y_i^k, (w^k)^T x_i) + \lambda \Omega(w^k),
$$
Norms, sparse norms, and computer vision

- Common norms include $p$-norm $\Omega(w) = \|w\|_p = (\sum_{i=1}^{p} w_i^p)^{1/p}$

$1$-norm promotes sparsity (prefers solutions with zero entries). Image denoising, total variation is useful, norm takes form:

$$\|w\|_1 = \sum_{i=2}^{N} |w_i|$$

Points of difference should be "sparse" (frequently zero). 

(Rodriguez, 2009)
Norms, sparse norms, and computer vision

- **Common norms include** $p$-norm $\Omega(w) = \|w\|_p = (\sum_{i=1}^{p} w_i^p)^{1/p}$
- **1-norm promotes sparsity** (prefer solutions with zero entries).

1-norm promotes sparsity (prefer solutions with zero entries).
Norms, sparse norms, and computer vision

- Common norms include $p$-norm $\Omega(w) = \|w\|_p = (\sum_{i=1}^{p} w_i^p)^{1/p}$
- 1-norm promotes sparsity (prefer solutions with zero entries).
- Image denoising, total variation is useful, norm takes form:

$$\Omega(w) = \sum_{i=2}^{N} |w_i - w_{i-1}|$$

(17.8)
Norms, sparse norms, and computer vision

- Common norms include $p$-norm $\Omega(w) = \|w\|_p = (\sum_{i=1}^{p} w_i^p)^{1/p}$
- 1-norm promotes sparsity (prefer solutions with zero entries).
- Image denoising, total variation is useful, norm takes form:

$$\Omega(w) = \sum_{i=2}^{N} |w_i - w_{i-1}|$$  \hspace{1cm} (17.8)

- Points of difference should be “sparse” (frequently zero).

(Rodriguez, 2009)
Submodular parameterization of a sparse convex norm

- Prefer convex norms since they can be solved.
Submodular parameterization of a sparse convex norm

- Prefer convex norms since they can be solved.
- For $w \in \mathbb{R}^V$, $\text{supp}(w) \in \{0,1\}^V$ has $\text{supp}(w)(v) = 1$ iff $w(v) > 0$
Submodular parameterization of a sparse convex norm

- Prefer convex norms since they can be solved.
- For $w \in \mathbb{R}^V$, $\text{supp}(w) \in \{0, 1\}^V$ has $\text{supp}(w)(v) = 1$ iff $w(v) > 0$
- Desirable sparse norm: count the non-zeros, $\|w\|_0 = 1^T \text{supp}(w)$. 

Ex: total variation is Lovász-ext. of graph cut, but many more!
Submodular parameterization of a sparse convex norm

- Prefer convex norms since they can be solved.
- For $w \in \mathbb{R}^V$, supp$(w) \in \{0, 1\}^V$ has supp$(w)(v) = 1$ iff $w(v) > 0$
- Desirable sparse norm: count the non-zeros, $\|w\|_0 = 1^T$ supp$(w)$.
- Using $\Omega(w) = \|w\|_0$ is NP-hard, instead we often optimize tightest convex relaxation, $\|w\|_1$ which is the convex envelope.
Submodular parameterization of a sparse convex norm

- Prefer convex norms since they can be solved.
- For $w \in \mathbb{R}^V$, supp($w$) $\in \{0, 1\}^V$ has supp($w$)($v$) = 1 iff $w(v) > 0$
- Desirable sparse norm: count the non-zeros, $\|w\|_0 = 1^T$ supp($w$).
- Using $\Omega(w) = \|w\|_0$ is NP-hard, instead we often optimize tightest convex relaxation, $\|w\|_1$ which is the convex envelope.
- With $\|w\|_0$ or its relaxation, each non-zero element has equal degree of penalty. Penalties do not interact.
Submodular parameterization of a sparse convex norm

- Prefer convex norms since they can be solved.
- For $w \in \mathbb{R}^V$, supp$(w) \in \{0, 1\}^V$ has supp$(w)(v) = 1$ iff $w(v) > 0$
- Desirable sparse norm: count the non-zeros, $\|w\|_0 = 1^T$ supp$(w)$.
- Using $\Omega(w) = \|w\|_0$ is NP-hard, instead we often optimize tightest convex relaxation, $\|w\|_1$ which is the convex envelope.
- With $\|w\|_0$ or its relaxation, each non-zero element has equal degree of penalty. Penalties do not interact.
- Given submodular function $f : 2^V \to \mathbb{R}_+$, $f$(supp$(w)$) measures the “complexity” of the non-zero pattern of $w$; can have more non-zero values if they cooperate (via $f$) with other non-zero values.
Submodular parameterization of a sparse convex norm

- Prefer convex norms since they can be solved.
- For $w \in \mathbb{R}^V$, $\text{supp}(w) \in \{0, 1\}^V$ has $\text{supp}(w)(v) = 1$ iff $w(v) > 0$.
- Desirable sparse norm: count the non-zeros, $\|w\|_0 = 1^T \text{supp}(w)$.
- Using $\Omega(w) = \|w\|_0$ is NP-hard, instead we often optimize tightest convex relaxation, $\|w\|_1$ which is the convex envelope.
- With $\|w\|_0$ or its relaxation, each non-zero element has equal degree of penalty. Penalties do not interact.
- Given submodular function $f : 2^V \rightarrow \mathbb{R}_+$, $f(\text{supp}(w))$ measures the "complexity" of the non-zero pattern of $w$; can have more non-zero values if they cooperate (via $f$) with other non-zero values.
- $f(\text{supp}(w))$ is hard to optimize, but it's convex envelope $\tilde{f}(\|w\|)$ (i.e., largest convex under-estimator of $f(\text{supp}(w))$) is obtained via the Lovász-extension $\tilde{f}$ of $f$ (Vondrák 2007, Bach 2010).
Submodular parameterization of a sparse convex norm

- Prefer convex norms since they can be solved.
- For $w \in \mathbb{R}^V$, $\text{supp}(w) \in \{0, 1\}^V$ has $\text{supp}(w)(v) = 1$ iff $w(v) > 0$
- Desirable sparse norm: count the non-zeros, $\|w\|_0 = 1^T \text{supp}(w)$.
- Using $\Omega(w) = \|w\|_0$ is NP-hard, instead we often optimize tightest convex relaxation, $\|w\|_1$ which is the convex envelope.
- With $\|w\|_0$ or its relaxation, each non-zero element has equal degree of penalty. Penalties do not interact.
- Given submodular function $f : 2^V \to \mathbb{R}_+$, $f(\text{supp}(w))$ measures the “complexity” of the non-zero pattern of $w$; can have more non-zero values if they cooperate (via $f$) with other non-zero values.
- $f(\text{supp}(w))$ is hard to optimize, but it's convex envelope $\tilde{f}(|w|)$ (i.e., largest convex under-estimator of $f(\text{supp}(w))$) is obtained via the Lovász-extension $\tilde{f}$ of $f$ (Vondrák 2007, Bach 2010).
- Submodular functions thus parameterize structured convex sparse norms via the Lovász-extension!
Submodular parameterization of a sparse convex norm

- Prefer convex norms since they can be solved.
- For $w \in \mathbb{R}^V$, $\text{supp}(w) \in \{0, 1\}^V$ has $\text{supp}(w)(v) = 1$ iff $w(v) > 0$
- Desirable sparse norm: count the non-zeros, $\|w\|_0 = 1^T \text{supp}(w)$.
- Using $\Omega(w) = \|w\|_0$ is NP-hard, instead we often optimize tightest convex relaxation, $\|w\|_1$ which is the convex envelope.
- With $\|w\|_0$ or its relaxation, each non-zero element has equal degree of penalty. Penalties do not interact.
- Given submodular function $f : 2^V \rightarrow \mathbb{R}_+$, $f(\text{supp}(w))$ measures the “complexity” of the non-zero pattern of $w$; can have more non-zero values if they cooperate (via $f$) with other non-zero values.
- $f(\text{supp}(w))$ is hard to optimize, but it’s convex envelope $\tilde{f}(|w|)$ (i.e., largest convex under-estimator of $f(\text{supp}(w))$) is obtained via the Lovász-extension $\tilde{f}$ of $f$ (Vondrák 2007, Bach 2010).
- Submodular functions thus parameterize structured convex sparse norms via the Lovász-extension!
- Ex: total variation is Lovász-ext. of graph cut, but $\exists$ many more!
Lovász extension and norms

- Using Lovász extension to define various norms of the form $\|w\|_{\tilde{f}} = \tilde{f}(|w|)$. This renders the function symmetric about all orthants (meaning, $\|w\|_{\tilde{f}} = \|b \odot w\|_{\tilde{f}}$ for any $b \in \{-1, 1\}^m$ and $\odot$ is element-wise multiplication).
Lovász extension and norms

- Using Lovász extension to define various norms of the form $\|w\|_{\tilde{f}} = \tilde{f}(|w|)$. This renders the function symmetric about all orthants (meaning, $\|w\|_{\tilde{f}} = \|b \odot w\|_{\tilde{f}}$ for any $b \in \{-1, 1\}^m$ and $\odot$ is element-wise multiplication).

- Simple example. The Lovász extension of the modular function $f(A) = |A|$ is the $\ell_1$ norm, and the Lovász extension of the modular function $f(A) = m(A)$ is the weighted $\ell_1$ norm.
Using Lovász extension to define various norms of the form $\|w\|_{\tilde{f}} = \tilde{f}(|w|)$. This renders the function symmetric about all orthants (meaning, $\|w\|_{\tilde{f}} = \|b \odot w\|_{\tilde{f}}$ for any $b \in \{-1, 1\}^m$ and $\odot$ is element-wise multiplication).

Simple example. The Lovász extension of the modular function $f(A) = |A|$ is the $\ell_1$ norm, and the Lovász extension of the modular function $f(A) = m(A)$ is the weighted $\ell_1$ norm.

With more general submodular functions, one can generate a large and interesting variety of norms, all of which have polyhedral contours (unlike, say, something like the $\ell_2$ norm).
Lovász extension and norms

- Using Lovász extension to define various norms of the form $\|w\|_{\tilde{f}} = \tilde{f}(|w|)$. This renders the function symmetric about all orthants (meaning, $\|w\|_{\tilde{f}} = \|b \odot w\|_{\tilde{f}}$ for any $b \in \{−1, 1\}^m$ and $\odot$ is element-wise multiplication).

- Simple example. The Lovász extension of the modular function $f(A) = |A|$ is the $\ell_1$ norm, and the Lovász extension of the modular function $f(A) = m(A)$ is the weighted $\ell_1$ norm.

- With more general submodular functions, one can generate a large and interesting variety of norms, all of which have polyhedral contours (unlike, say, something like the $\ell_2$ norm).

- Hence, not all norms come from the Lovász extension of some submodular function.
Lovász extension and norms

- Using Lovász extension to define various norms of the form $\|w\|_{\tilde{f}} = \tilde{f}(|w|)$. This renders the function symmetric about all orthants (meaning, $\|w\|_{\tilde{f}} = \|b \odot w\|_{\tilde{f}}$ for any $b \in \{-1, 1\}^m$ and $\odot$ is element-wise multiplication).

- Simple example. The Lovász extension of the modular function $f(A) = |A|$ is the $\ell_1$ norm, and the Lovász extension of the modular function $f(A) = m(A)$ is the weighted $\ell_1$ norm.

- With more general submodular functions, one can generate a large and interesting variety of norms, all of which have polyhedral contours (unlike, say, something like the $\ell_2$ norm).

- Hence, not all norms come from the Lovász extension of some submodular function.

- Similarly, not all convex functions are the Lovász extension of some submodular function.
Lovász extension and norms

- Using Lovász extension to define various norms of the form \( \|w\|_{\tilde{f}} = \tilde{f}(\|w\|) \). This renders the function symmetric about all orthants (meaning, \( \|w\|_{\tilde{f}} = \|b \odot w\|_{\tilde{f}} \) for any \( b \in \{-1, 1\}^m \) and \( \odot \) is element-wise multiplication).

- Simple example. The Lovász extension of the modular function \( f(A) = |A| \) is the \( \ell_1 \) norm, and the Lovász extension of the modular function \( f(A) = m(A) \) is the weighted \( \ell_1 \) norm.

- With more general submodular functions, one can generate a large and interesting variety of norms, all of which have polyhedral contours (unlike, say, something like the \( \ell_2 \) norm).

- Hence, not all norms come from the Lovász extension of some submodular function.

- Similarly, not all convex functions are the Lovász extension of some submodular function.

- Bach-2011 has a complete discussion of this.
The \textit{concave} closure is defined as:

$$\hat{f}(x) = \max_{p \in \triangle^n(x)} \sum_{S \subseteq V} p_S f(S)$$

(17.9)

where $\triangle^n(x) =$

$$\left\{ p \in \mathbb{R}^{2^n} : \sum_{S \subseteq V} p_S = 1, p_S \geq 0 \forall S \subseteq V, \text{ and } \sum_{S \subseteq V} p_S 1_S = x \right\}$$
The concave closure is defined as:

\[
\hat{f}(x) = \max_{p \in \Delta^n(x)} \sum_{S \subseteq V} p_S f(S)
\]  

(17.9)

where \( \Delta^n(x) = \{ p \in \mathbb{R}^{2^n} : \sum_{S \subseteq V} p_S = 1, p_S \geq 0 \forall S \subseteq V, \& \sum_{S \subseteq V} p_S 1_S = x \} \)

This is tight at the hypercube vertices, concave, and the concave envelope for the dual reasons as the convex closure.
The concave closure is defined as:

$$\hat{f}(x) = \max_{p \in \Delta^n(x)} \sum_{S \subseteq V} p_S f(S)$$  \hspace{1cm} (17.9)

where $\Delta^n(x) = \left\{ p \in \mathbb{R}^{2^n} : \sum_{S \subseteq V} p_S = 1, p_S \geq 0 \forall S \subseteq V, \ & \sum_{S \subseteq V} p_S 1_S = x \right\}$

This is tight at the hypercube vertices, concave, and the concave envelope for the dual reasons as the convex closure.

Unlike the convex extension, the concave closure is defined by the Lovász extension iff $f$ is a supermodular function.
The **concave** closure is defined as:

\[
\hat{f}(x) = \max_{p \in \triangle^n(x)} \sum_{S \subseteq V} p_S f(S)
\]

where \(\triangle^n(x) = \left\{ p \in \mathbb{R}^{2^n} : \sum_{S \subseteq V} p_S = 1, p_S \geq 0 \forall S \subseteq V, \sum_{S \subseteq V} p_S \mathbf{1}_S = x \right\} \)

- This is tight at the hypercube vertices, concave, and the concave envelope for the dual reasons as the convex closure.
- Unlike the convex extension, the concave closure is defined by the Lovász extension iff \(f\) is a supermodular function.
- When \(f\) is submodular, even evaluating \(\hat{f}\) is NP-hard (rough intuition: submodular maximization is NP-hard (reduction to set cover), if we could evaluate \(\hat{f}\) in poly time, we can maximize concave function to solve submodular maximization in poly time).
Rather than the concave closure, multi-linear extension is used as a surrogate. For $x \in [0, 1]^V = [0, 1]^n$

$$\tilde{f}(x) = \sum_{S \subseteq V} f(S) \prod_{i \in S} x_i \prod_{i \in V \setminus S} (1 - x_i) = E_{S \sim x}[f(S)]$$  (17.10)

What to do?

1) **Multilinear extension**

2) **Restricted class of submodular function** that has easier concave closure

3) **Polynomial relaxations**
Multilinear extension

Rather than the concave closure, multi-linear extension is used as a surrogate. For \( x \in [0, 1]^V = [0, 1]^n \)

\[
\tilde{f}(x) = \sum_{S \subseteq V} f(S) \prod_{i \in S} x_i \prod_{i \in V \setminus S} (1 - x_i) = E_{S \sim x}[f(S)]
\]  

(17.10)

Can be viewed as expected value of \( f(S) \) where \( S \) is a random set distributed via \( x \), so \( \Pr(v \in S) = x_v \) and is independent of \( \Pr(u \in S) = x_u, \quad v \neq u \).

\[ x \in \{0, 1\}^V \]
Multilinear extension

- Rather than the concave closure, multi-linear extension is used as a surrogate. For $x \in [0, 1]^V = [0, 1]^n$

$$\tilde{f}(x) = \sum_{S \subseteq V} f(S) \prod_{i \in S} x_i \prod_{i \in V \setminus S} (1 - x_i) = E_{S \sim x}[f(S)] \quad (17.10)$$

- Can be viewed as expected value of $f(S')$ where $S$ is a random set distributed via $x$, so $\Pr(v \in S) = x_v$ and is independent of $\Pr(u \in S) = x_u$, $v \neq u$.

- This is tight at the hypercube vertices (immediate, since $f(1_A)$ yields only one term in the sum non-zero, namely the one where $S = A$).
Multilinear extension

- Rather than the concave closure, multi-linear extension is used as a surrogate. For $x \in [0, 1]^V = [0, 1]^n$

$$\tilde{f}(x) = \sum_{S \subseteq V} f(S) \prod_{i \in S} x_i \prod_{i \in V \setminus S} (1 - x_i) = E_{S \sim x}[f(S')] \quad (17.10)$$

- Can be viewed as expected value of $f(S')$ where $S$ is a random set distributed via $x$, so $\Pr(v \in S) = x_v$ and is independent of $\Pr(u \in S) = x_u$, $v \neq u$.

- This is tight at the hypercube vertices (immediate, since $f(1_A)$ yields only one term in the sum non-zero, namely the one where $S = A$).

- Why called multilinear (multi-linear) extension? It is linear in each of its arguments (i.e., $\tilde{f}(x_1, x_2, \ldots, \alpha x_k + \beta x'_k, \ldots, x_n) = \alpha \tilde{f}(x_1, x_2, \ldots, x_k, \ldots, x_n) + \beta \tilde{f}(x_1, x_2, \ldots, x'_k, \ldots, x_n)$
Rather than the concave closure, multi-linear extension is used as a surrogate. For $x \in [0, 1]^V = [0, 1]^n$,

$$\tilde{f}(x) = \sum_{S \subseteq V} f(S) \prod_{i \in S} x_i \prod_{i \in V \setminus S} (1 - x_i) = E_{S \sim x}[f(S')]$$  \hspace{1cm} (17.10)

Can be viewed as expected value of $f(S')$ where $S$ is a random set distributed via $x$, so $\Pr(v \in S) = x_v$ and is independent of $\Pr(u \in S) = x_u, v \neq u$.

This is tight at the hypercube vertices (immediate, since $f(1_A)$ yields only one term in the sum non-zero, namely the one where $S = A$).

Why called multilinear (multi-linear) extension? It is linear in each of its arguments (i.e., $\tilde{f}(x_1, x_2, \ldots, \alpha x_k + \beta x'_k, \ldots, x_n) = \alpha \tilde{f}(x_1, x_2, \ldots, x_k, \ldots, x_n) + \beta \tilde{f}(x_1, x_2, \ldots, x'_k, \ldots, x_n)$)

This is unfortunately not concave. However there are some useful properties.
Lemma 17.4.1

Let \( \tilde{f}(\tilde{x}) \) be the multilinear extension of a set function \( f : 2^V \rightarrow \mathbb{R} \). Then:

- If \( f \) is monotone non-decreasing, then \( \frac{\partial \tilde{f}}{\partial x_v} \geq 0 \) for all \( v \in V \) within \([0, 1]^V\) (i.e., \( \tilde{f} \) is also monotone non-decreasing).
- If \( f \) is submodular, then \( \tilde{f} \) has an antitone supergradient, i.e.,
  \[
  \frac{\partial^2 \tilde{f}}{\partial x_i \partial x_j} \leq 0 \quad \text{for all } i, j \in V \text{ within } [0, 1]^V.
  \]

Proof.

- First part (monotonicity). Choose \( x \in [0, 1]^V \) and let \( S \sim x \) be random where \( x \) is treated as a distribution (so elements \( v \) is chosen with probability \( x_v \) independently of any other element).
Since \( \tilde{f} \) is multilinear, derivative is a simple difference when only one argument varies, i.e.,

\[
\frac{\partial \tilde{f}}{\partial x_v} = \tilde{f}(x_1, x_2, \ldots, x_{v_1}, 1, x_{v+1}, \ldots, x_n) - \tilde{f}(x_1, x_2, \ldots, x_{v_1}, 0, x_{v+1}, \ldots, x_n)
\]

\[= ES_{x \sim x}[f(S + v)] - ES_{x \sim x}[f(S - v)]\]  

where the final part follows due to monotonicity of each argument, i.e.,

\( f(S + i) \geq f(S - i) \) for any \( S \) and \( i \in V \).
Second part of proof (antitone supergradient) also relies on simple consequence of multilinearity, namely multilinearity of the derivative as well. In this case
\[
\frac{\partial^2 \tilde{f}}{\partial x_i \partial x_j} = \frac{\partial \tilde{f}}{\partial x_j}(x_1, \ldots, x_{i-1}, 1, x_{i+1}, \ldots, x_n) \tag{17.15}
\]
\[
- \frac{\partial \tilde{f}}{\partial x_j}(x_1, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_n) \tag{17.16}
\]
\[
= E_{S \sim x}[f(S + i + j) - f(S + i - j)] \tag{17.17}
\]
\[
- E_{S \sim x}[f(S - i + j) - f(S - i - j)] \tag{17.18}
\]
\[
\leq 0 \tag{17.19}
\]
since by submodularity, we have
\[
f(S + i - j) + f(S - i + j) \geq f(S + i + j) + f(S - i - j) \tag{17.20}
\]
Corollary 17.4.2

let \( f \) be a function and \( \tilde{f} \) its multilinear extension on \([0, 1]^V\).

- if \( f \) is monotone non-decreasing then \( \tilde{f} \) is non-decreasing along any strictly non-negative direction (i.e., \( \tilde{f}(x) \leq \tilde{f}(y) \) whenever \( x \leq y \), or \( \tilde{f}(x) \leq \tilde{f}(x + \epsilon \mathbf{1}_v) \) for any \( v \in V \) and any \( \epsilon \geq 0 \).
- If \( f \) is submodular, then \( \tilde{f} \) is concave along any non-negative direction (i.e., the function \( g(\alpha) = \tilde{f}(x + \alpha z) \) is 1-D concave in \( \alpha \) for any \( z \in \mathbb{R}_+ \)).
- If \( f \) is submodular than \( \tilde{f} \) is convex along any diagonal direction (i.e., the function \( g(\alpha) = \tilde{f}(x + \alpha(\mathbf{1}_v - \mathbf{1}_u)) \) is 1-D convex in \( \alpha \) for any \( u \neq v \).
We’ve spent much time discussing SFM and the polymatroidal polytope, and in general polyhedral approaches for SFM.
We’ve spent much time discussing SFM and the polymatroidal polytope, and in general polyhedral approaches for SFM.

Most of the approaches for submodular max have not used such an approach, probably due to the difficulty in computing the “concave extension” of a submodular function (the convex extension is easy, namely the Lovász extension).
We’ve spent much time discussing SFM and the polymatroidal polytope, and in general polyhedral approaches for SFM.

Most of the approaches for submodular max have not used such an approach, probably due to the difficulty in computing the “concave extension” of a submodular function (the convex extension is easy, namely the Lovász extension).

A paper by Chekuri, Vondrak, and Zenklusen (2011) make some progress on this front using multilinear extensions.
Multilinear extension (review)

Definition 17.5.1

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

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

(17.21)

- Note that $F(x) = Ef(\hat{x})$ where $\hat{x}$ is a random binary vector over $\{0, 1\}^V$ with elements independent w. probability $x_i$ for $\hat{x}_i$.
- While this is defined for any set function, we have:

Lemma 17.5.2

Let $F : [0, 1]^V \to \mathbb{R}$ be multilinear extension of set function $f : 2^V \to \mathbb{R}$, then

- If $f$ is monotone non-decreasing, then $\frac{\partial F}{\partial x_i} \geq 0$ for all $i \in V, x \in [0, 1]^V$.
- If $f$ is submodular, then $\frac{\partial^2 F}{\partial x_i \partial x_j} \leq 0$ for all $i, j \in V, x \in [0, 1]^V$. 
Basic idea: Given a set of constraints $\mathcal{I}$, we form a polytope $P_\mathcal{I}$ such that $\{1_I : I \in \mathcal{I}\} \subseteq P_\mathcal{I}$

We find $\max_{x \in P_\mathcal{I}} F(x)$ where $F(x)$ is the multi-linear extension of $f$, to find a fractional solution $x^*$

We then round $x^*$ to a point on the hypercube, thus giving us a solution to the discrete problem.
In the recent paper by Chekuri, Vondrak, and Zenklusen, they show:

1) constant factor approximation algorithm for $\max_{x \in P} F(x)$ for any down-monotone solvable polytope $P$ and $F$ multilinear extension of any non-negative submodular function.

2) A randomized rounding (pipage rounding) scheme to obtain an integer solution.

3) An optimal $(1 - 1/e)$ instance of their rounding scheme that can be used for a variety of interesting independence systems, including $O(1)$ knapsacks, $k$ matroids and $O(1)$ knapsacks, a $k$-matchoid and `sparse packing integer programs, and unsplittable flow in paths and trees. Also, Vondrak showed that this scheme achieves the $1/e$ curvature based bound for any matroid, which matches the bound we had earlier for uniform matroids with standard greedy.

In general, one needs to do Monte-Carlo methods to estimate the multilinear extension (so further approximations would apply).
In the recent paper by Chekuri, Vondrak, and Zenklusen, they show:

1) constant factor approximation algorithm for $\max \{ F(x) : x \in P \}$ for any down-monotone solvable polytope $P$ and $F$ multilinear extension of any non-negative submodular function.
In the recent paper by Chekuri, Vondrak, and Zenklusen, they show:

1) constant factor approximation algorithm for $\max \{ F(x) : x \in P \}$ for any down-monotone solvable polytope $P$ and $F$ multilinear extension of any non-negative submodular function.

2) A randomized rounding (pipage rounding) scheme to obtain an integer solution.
In the recent paper by Chekuri, Vondrak, and Zenklusen, they show:

1) constant factor approximation algorithm for $\max \{ F(x) : x \in P \}$ for any down-monotone solvable polytope $P$ and $F$ multilinear extension of any non-negative submodular function.

2) A randomized rounding (pipage rounding) scheme to obtain an integer solution

3) An optimal $(1 - 1/e)$ instance of their rounding scheme that can be used for a variety of interesting independence systems, including $O(1)$ knapsacks, $k$ matroids and $O(1)$ knapsacks, a $k$-matchoid and $\ell$ sparse packing integer programs, and unsplittable flow in paths and trees.
In the recent paper by Chekuri, Vondrak, and Zenklusen, they show:

1) constant factor approximation algorithm for $\max \{F(x) : x \in P\}$ for any down-monotone solvable polytope $P$ and $F$ multilinear extension of any non-negative submodular function.

2) A randomized rounding (pipage rounding) scheme to obtain an integer solution

3) An optimal $(1 - 1/e)$ instance of their rounding scheme that can be used for a variety of interesting independence systems, including $O(1)$ knapsacks, $k$ matroids and $O(1)$ knapsacks, a $k$-matchoid and $\ell$ sparse packing integer programs, and unsplittable flow in paths and trees.

Also, Vondrak showed that this scheme achieves the curvature based bound for any matroid, which matches the bound we had earlier for uniform matroids with standard greedy.
In the recent paper by Chekuri, Vondrak, and Zenklusen, they show:

1) constant factor approximation algorithm for $\max \{ F(x) : x \in P \}$ for any down-monotone solvable polytope $P$ and $F$ multilinear extension of any non-negative submodular function.

2) A randomized rounding (pipage rounding) scheme to obtain an integer solution

3) An optimal $(1 - 1/e)$ instance of their rounding scheme that can be used for a variety of interesting independence systems, including $O(1)$ knapsacks, $k$ matroids and $O(1)$ knapsacks, a $k$-matchoid and $\ell$ sparse packing integer programs, and unsplittable flow in paths and trees.

Also, Vondrak showed that this scheme achieves the $\frac{1}{c}(1 - e^{-c})$ curvature based bound for any matroid, which matches the bound we had earlier for uniform matroids with standard greedy.

In general, one needs to do Monte-Carlo methods to estimate the multilinear extension (so further approximations would apply).
The next slide comes from lecture 10.
A polymatroid function’s polyhedron is a polymatroid.

**Theorem 17.6.1**

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

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

$$= \min \left( x(A) + f(E \setminus A) : A \subseteq E \right) \quad (17.10)$$

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:

$$\text{rank} \left( \frac{1}{\epsilon} 1_{E \setminus B} \right) = f(E \setminus B) = \max \left\{ y(E \setminus B) : y \in P_{f}^{+} \right\} \quad (17.11)$$

In fact, we will ultimately see a number of important consequences of this theorem (other than just that $P_{f}^{+}$ is a polymatroid).
The next slide comes from lecture 11.
• Considering Theorem ??, the matroid case is now a special case, where we have that:

**Corollary 17.6.2**

We have that:

\[
\max \{ y(E) : y \in P_{\text{ind. set}}(M), y \leq x \} = \min \{ r_M(A) + x(E \setminus A) : A \subseteq E \}
\]

\[(17.21)\]

where \( r_M \) is the matroid rank function of some matroid.
Consider

\[ P_r^+ = \{ x \in \mathbb{R}^E : x \geq 0, x(A) \leq r_M(A), \forall A \subseteq E \} \]  

(17.22)
Consider

\[ P_r^+ = \{ x \in \mathbb{R}^E : x \geq 0, x(A) \leq r_M(A), \forall A \subseteq E \} \]  \hspace{1cm} (17.22)

Suppose we have any \( x \in \mathbb{R}_+^E \) such that \( x \not\in P_r^+ \).
Most violated inequality problem in matroid polytope case

- Consider

\[ P_r^+ = \{ x \in \mathbb{R}^E : x \geq 0, x(A) \leq r_M(A), \forall A \subseteq E \} \quad (17.22) \]

- Suppose we have any \( x \in \mathbb{R}_+^E \) such that \( x \not\in P_r^+ \).

- 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 \( A \in \mathcal{W} \).
Consider

\[ P_r^+ = \{ x \in \mathbb{R}^E : x \geq 0, x(A) \leq r_M(A), \forall A \subseteq E \} \]  

(17.22)

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 valuated as:

\[
\max \{ x(A) - r_M(A) : A \in \mathcal{W} \} = \max \{ x(A) - r_M(A) : A \subseteq E \} \]  

(17.23)
Most violated inequality problem in matroid polytope case

- Consider

\[ P_r^+ = \{ x \in \mathbb{R}^E : x \geq 0, x(A) \leq r_M(A), \forall A \subseteq E \} \tag{17.22} \]

- Suppose we have any \( x \in \mathbb{R}^E_+ \) such that \( x \not\in P_r^+ \).

- 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 valuated as:

\[ \max \{ x(A) - r_M(A) : A \in \mathcal{W} \} = \max \{ x(A) - r_M(A) : A \subseteq E \} \tag{17.23} \]

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

\[ \min \{ r_M(A) + x(E \setminus A) : A \subseteq E \} \tag{17.24} \]
Consider

$$P_f^+ = \{ x \in \mathbb{R}^E : x \geq 0, x(A) \leq f(A), \forall A \subseteq E \} \quad (17.25)$$
Consider

\[ P_f^+ = \{ x \in \mathbb{R}^E : x \geq 0, x(A) \leq f(A), \forall A \subseteq E \} \]  \hspace{1cm} (17.25)

Suppose we have any \( x \in \mathbb{R}^E_+ \) such that \( x \not\in P_f^+ \).
Consider

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

(17.25)

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 \( A \in \mathcal{W} \).
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:

$$\max \{x(A) - f(A) : A \in \mathcal{W}\} = \max \{x(A) - f(A) : A \subseteq E\} \quad (17.26)$$
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:

$$\max \{x(A) - f(A) : A \in \mathcal{W}\} = \max \{x(A) - f(A) : A \subseteq E\} \quad (17.26)$$

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

$$\min \{f(A) + x(E \setminus A) : A \subseteq E\} \quad (17.27)$$
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:

$$\max \{x(A) - f(A) : A \in \mathcal{W}\} = \max \{x(A) - f(A) : A \subseteq E\} \quad (17.26)$$

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

$$\min \{f(A) + x(E \setminus A) : A \subseteq E\} \quad (17.27)$$

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).
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 valued as:

$$\max \{x(A) - f(A) : A \in \mathcal{W}\} = \max \{x(A) - f(A) : A \subseteq E\} \quad (17.26)$$

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

$$\min \{f(A) + x(E \setminus A) : A \subseteq E\} \quad (17.27)$$

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 ultimately answer how general this form of SFM is.
The following three slides are review from lecture 6.
Definition 17.7.3 (closed/flat/subspace)

A subset $A \subseteq E$ is closed (equivalently, a flat or a subspace) of matroid $M$ if for all $x \in E \setminus A$, $r(A \cup \{x\}) = r(A) + 1$.

Definition: A hyperplane is a flat of rank $r(M) - 1$.

Definition 17.7.4 (closure)

Given $A \subseteq E$, the closure (or span) of $A$, is defined by

$$\text{span}(A) = \{b \in E : r(A \cup \{b\}) = r(A)\}.$$ 

Therefore, a closed set $A$ has $\text{span}(A) = A$.

Definition 17.7.5 (circuit)

A subset $A \subseteq E$ is circuit or a cycle if it is an inclusionwise-minimal dependent set (i.e., if $r(A) < |A|$ and for any $a \in A$, $r(A \setminus \{a\}) = |A| - 1$).
A set is independent if and only if it contains no circuit. Therefore, it is not surprising that circuits can also characterize a matroid.

**Theorem 17.7.3 (Matroid by circuits)**

Let $E$ be a set and $\mathcal{C}$ be a collection of subsets of $E$ that satisfy the following three properties:

1. **(C1):** $\emptyset \notin \mathcal{C}$
2. **(C2):** if $C_1, C_2 \in \mathcal{C}$ and $C_1 \subseteq C_2$, then $C_1 = C_2$.
3. **(C3):** if $C_1, C_2 \in \mathcal{C}$ with $C_1 \neq C_2$, and $e \in C_1 \cap C_2$, then there exists a $C_3 \in \mathcal{C}$ such that $C_3 \subseteq (C_1 \cup C_2) \setminus \{e\}$. 
Matroids by circuits

Several circuit definitions for matroids.

**Theorem 17.7.3 (Matroid by circuits)**

Let $E$ be a set and $C$ be a collection of nonempty subsets of $E$, such that no two sets in $C$ are contained in each other. Then the following are equivalent.

1. $C$ is the collection of circuits of a matroid;
2. if $C, C' \in C$, and $x \in C \cap C'$, then $(C \cup C') \setminus \{x\}$ contains a set in $C$;
3. if $C, C' \in C$, and $x \in C \cap C'$, and $y \in C \setminus C'$, then $(C \cup C') \setminus \{x\}$ contains a set in $C$ containing $y$;

Again, think about this for a moment in terms of linear spaces and matrices, and spanning trees.
Lemma 17.7.1

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 $C_1 \cup C_2 \in I \cup \{e\}$.

Then $e \in C_1 \setminus C_2$, and by Lemma (C2), there is a circuit $C_3$ of $M$ such that $C_3 \in (C_1 \cup C_2) \cap \{e\}$.

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 $e$).
Lemma 17.7.1

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 $C_1 \cup C_2 \subseteq I \cup \{e\}$. 

---

Prof. Jeff Bilmes  
EE563/Spring 2018/Submodularity - Lecture 17 - May 23st, 2018  
F38/54 (pg.92/192)
Lemma 17.7.1

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. $C_3 \subseteq (C_1 \cup C_2) \setminus \{e\} \subseteq I$.
Lemma 17.7.1

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 $M$ such that $C_3 \subseteq (C_1 \cup C_2) \setminus \{e\} \subseteq I$.
- This contradicts the independence of $I$. 

[Box]
**Lemma 17.7.1**

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 $e$).
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).
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).
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.
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.
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 \( C(I, e) = \emptyset \), since no circuit is created in this case.
Lemma 17.7.2

Let $\mathcal{B}(D)$ be the set of bases of any set $D$. Then, given matroid $M = (E, I)$, and any loop-free (i.e., no dependent singleton elements) set $D \subseteq E$, we have:

$$\bigcup_{B \in \mathcal{B}(D)} B = D. \quad (17.28)$$
Lemma 17.7.2

Let $\mathcal{B}(D)$ be the set of bases of any set $D$. Then, given matroid $\mathcal{M} = (E, I)$, and any loop-free (i.e., no dependent singleton elements) set $D \subseteq E$, we have:

$$\bigcup_{B \in \mathcal{B}(D)} B = D.$$  \hspace{1cm} (17.28)

Proof.

1. Define $D' \triangleq \bigcup_{B \in \mathcal{B}(D)} \subseteq D$, suppose $\exists d \in D$ such that $d \notin D'$.
Lemma 17.7.2

Let $\mathcal{B}(D)$ be the set of bases of any set $D$. Then, given matroid $\mathcal{M} = (E, \mathcal{I})$, and any loop-free (i.e., no dependent singleton elements) set $D \subseteq E$, we have:

$$\bigcup_{B \in \mathcal{B}(D)} B = D. \quad (17.28)$$

Proof.

- Define $D' \triangleq \bigcup_{B \in \mathcal{B}(D)} B \subseteq D$, suppose $\exists d \in D$ such that $d \notin D'$.
- Hence, $\forall B \in \mathcal{B}(D)$ we have $d \notin B$, and $B + d$ must contain a single circuit for any $B$, namely $C(B, d)$. 
Union of matroid bases of a set

Lemma 17.7.2

Let $\mathcal{B}(D)$ be the set of bases of any set $D$. Then, given matroid $\mathcal{M} = (E, \mathcal{I})$, and any loop-free (i.e., no dependent singleton elements) set $D \subseteq E$, we have:

$$
\bigcup_{B \in \mathcal{B}(D)} B = D. \tag{17.28}
$$

Proof.

- Define $D' \triangleq \bigcup_{B \in \mathcal{B}(D)} \subseteq D$, suppose $\exists d \in D$ such that $d \notin D'$.
- Hence, $\forall B \in \mathcal{B}(D)$ we have $d \notin B$, and $B + d$ must contain a single circuit for any $B$, namely $C(B, d)$.
- Then choose $d' \in C(B, d)$ with $d' \neq d$. 

Prof. Jeff Bilmes
Lemma 17.7.2

Let $\mathcal{B}(D)$ be the set of bases of any set $D$. Then, given matroid $\mathcal{M} = (E, \mathcal{I})$, and any loop-free (i.e., no dependent singleton elements) set $D \subseteq E$, we have:

$$\bigcup_{B \in \mathcal{B}(D)} B = D. \quad (17.28)$$

Proof.

- Define $D' \triangleq \bigcup_{B \in \mathcal{B}(D)} \subseteq D$, suppose $\exists d \in D$ such that $d \notin D'$.
- Hence, $\forall B \in \mathcal{B}(D)$ we have $d \notin B$, and $B + d$ must contain a single circuit for any $B$, namely $C(B, d)$.
- Then choose $d' \in C(B, d)$ with $d' \neq d$.
- Then $B + d - d'$ is independent size-$|B|$ subset of $D$ and hence spans $D$, and thus is a $d$-containing member of $\mathcal{B}(D)$, contradicting $d \notin D'$. 

Prof. Jeff Bilmes
The $\text{sat}$ function $\equiv$ Polymatroid Closure

- Thus, in a matroid, closure (span) of a set $A$ are all items that $A$ spans (eq. that depend on $A$).
The \textit{sat} function $\equiv$ 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.
The \textit{sat} function $=$ 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 $f$. 
The \textit{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 $x \in Pf$ for polymatroid function $f$.
- Again, recall, tight sets are closed under union and intersection, and therefore form a distributive lattice.
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.
The \textit{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

$$D(x) = \{ A : A \subseteq E, x(A) = f(A) \}$$  \hspace{1cm} (17.29)
The sat function $\equiv$ Polymatroid Closure

Now given $x \in P_f^+$:

$$D(x) = \{ A : A \subseteq E, x(A) = f(A) \} \quad (17.30)$$

$$= \{ A : f(A) - x(A) = 0 \} \quad (17.31)$$
The \textit{sat} function $= \text{Polymatroid Closure}$

- Now given $x \in P_f^+$:

\[
\mathcal{D}(x) = \{ A : A \subseteq E, x(A) = f(A) \} \quad (17.30)
\]

\[
= \{ A : f(A) - x(A) = 0 \} \quad (17.31)
\]

- Since $x \in P_f^+$ and $f$ is presumed to be polymatroid function, we see $f'(A) = f(A) - x(A)$ is a non-negative submodular function, and $\mathcal{D}(x)$ are the zero-valued minimizers (if any) of $f'(A)$. 
The sat function = Polymatroid Closure

- Now given $x \in P^+_f$:

$$D(x) = \{ A : A \subseteq E, x(A) = f(A) \}$$

$$= \{ A : f(A) - x(A) = 0 \}$$

(17.30) (17.31)

- Since $x \in P^+_f$ and $f$ is presumed to be polymatroid function, we see $f'(A) = f(A) - x(A)$ is a non-negative submodular function, and $D(x)$ are the zero-valued minimizers (if any) of $f'(A)$.

- The zero-valued minimizers of $f'$ are thus closed under union and intersection.
Now given $x \in P_f^+$:

$$D(x) = \{ A : A \subseteq E, x(A) = f(A) \}$$

$$= \{ A : f(A) - x(A) = 0 \}$$

Since $x \in P_f^+$ and $f$ is presumed to be polymatroid function, we see $f'(A) = f(A) - x(A)$ is a non-negative submodular function, and $D(x)$ are the zero-valued minimizers (if any) of $f'(A)$.

The zero-valued minimizers of $f'$ are thus closed under union and intersection.

In fact, this is true for all minimizers of a submodular function as stated in the next theorem.
Minimizers of a Submodular Function form a lattice

**Theorem 17.8.1**

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

Proof.

Since $A$ and $B$ are minimizers, we have $f(A) = f(B)$ and $f(A \setminus B) = f(A \cup B)$. By submodularity, we have

$$f(A) + f(B) = f(A \cup B) + f(A \setminus B)$$

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

Thus, the minimizers of a submodular function form a lattice, and there is a maximal and a minimal minimizer of every submodular function.
Minimizers of a Submodular Function form a lattice

**Theorem 17.8.1**

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

**Proof.**
Minimizers of a Submodular Function form a lattice

**Theorem 17.8.1**

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

**Proof.**

Since $A$ and $B$ are minimizers, we have $f(A) = f(B) \leq f(A \cap B)$ and $f(A) = f(B) \leq f(A \cup B)$. 
Minimizers of a Submodular Function form a lattice

**Theorem 17.8.1**

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

**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

\[
f(A) + f(B) \geq f(A \cup B) + f(A \cap B) \quad (17.32)
\]
Theorem 17.8.1

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

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

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B) \quad (17.32)$$

Hence, we must have $f(A) = f(B) = f(A \cup B) = f(A \cap B)$.
Minimizers of a Submodular Function form a lattice

**Theorem 17.8.1**

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

**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

\[
f(A) + f(B) \geq f(A \cup B) + f(A \cap B) \quad (17.32)
\]

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.
Matroid closure is generalized by the unique maximal element in $D(x)$, also called the polymatroid closure or \textit{sat} (saturation function).
The \textit{sat} function $= \text{Polymatroid Closure}$

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

$$\text{cl}(x) \overset{\text{def}}{=} \text{sat}(x) \overset{\text{def}}{=} \bigcup \{ A : A \in \mathcal{D}(x) \}$$  \hspace{1cm} (17.33)
The \textit{sat} function $\equiv$ Polymatroid Closure

- Matroid closure is generalized by the unique maximal element in $\mathcal{D}(x)$, also called the polymatroid closure or \textit{sat} (saturation function).

- For some $x \in P_f$, we have defined:

\[
\text{cl}(x) \overset{\text{def}}{=} \text{sat}(x) \overset{\text{def}}{=} \bigcup \{ A : A \in \mathcal{D}(x) \} \\
= \bigcup \{ A : A \subseteq E, x(A) = f(A) \} \quad (17.33)
\]
The sat function $\equiv$ Polymatroid Closure

- Matroid closure is generalized by the unique maximal element in $\mathcal{D}(x)$, also called the polymatroid closure or sat (saturation function).

- For some $x \in P_f$, we have defined:

$$
\text{cl}(x) \overset{\text{def}}{=} \text{sat}(x) \overset{\text{def}}{=} \bigcup \{ A : A \in \mathcal{D}(x) \}
$$

$$
= \bigcup \{ A : A \subseteq E, x(A) = f(A) \}
$$

$$
= \{ e : e \in E, \forall \alpha > 0, x + \alpha \mathbf{1}_e \notin P_f \}
$$
The \( \text{sat} \) function \( = \) Polymatroid Closure

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

\[
\text{cl}(x) \overset{\text{def}}{=} \text{sat}(x) \overset{\text{def}}{=} \bigcup \{A : A \in \mathcal{D}(x)\} \\
= \bigcup \{A : A \subseteq E, x(A) = f(A)\} \\
= \{e : e \in E, \forall \alpha > 0, x + \alpha 1_e \notin P_f\}
\]

Hence, \( \text{sat}(x) \) is the maximal (zero-valued) minimizer of the submodular function \( f_x(A) \overset{\Delta}{=} f(A) - x(A) \).
The sat function $\text{sat}$ function $= \text{Polymatroid Closure}$

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

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

Hence, sat$(x)$ is the maximal (zero-valued) minimizer of the submodular function $f_x(A) \overset{\Delta}{=} f(A) - x(A)$.

Eq. (17.35) says that sat consists of elements of 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.
The sat function $= \text{Polymatroid Closure}$

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

\[
\text{cl}(x) \overset{\text{def}}{=} \text{sat}(x) \overset{\text{def}}{=} \bigcup \{ A : A \in \mathcal{D}(x) \} \\
= \bigcup \{ A : A \subseteq E, x(A) = f(A) \} \\
= \{ e : e \in E, \forall \alpha > 0, x + \alpha 1_e \notin P_f \}
\]  

(17.33)  
(17.34)  
(17.35)

- Hence, $\text{sat}(x)$ is the maximal (zero-valued) minimizer of the submodular function $f_x(A) \overset{\triangle}{=} f(A) - x(A)$.
- Eq. (17.35) says that sat consists of elements of 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.
Consider matroid \((E, \mathcal{I}) = (E, r)\), some \(I \in \mathcal{I}\). Then \(1_I \in \mathcal{P}_r\) and
\[
\begin{align*}
\mathcal{D}(1_I) &= \{ A : 1_I(A) = r(A) \} \\
\end{align*}
\] (17.36)
Consider matroid \((E, I) = (E, r)\), some \(I \in \mathcal{I}\). Then \(1_I \in P_r\) and

\[
D(1_I) = \{ A : 1_I(A) = r(A) \}
\]

and

\[
sat(1_I)
\]
Consider matroid \((E, \mathcal{I}) = (E, r)\), some \(I \in \mathcal{I}\). Then \(1_I \in P_r\) and

\[
\mathcal{D}(1_I) = \{ A : 1_I(A) = r(A) \} \quad (17.36)
\]

and

\[
\text{sat}(1_I) = \bigcup \{ A : A \subseteq E, A \in \mathcal{D}(1_I) \} \quad (17.37)
\]

Notice that \(1_I(A) = |I \setminus A| \geq |I|\). Intuitively, consider an \(A \in \mathcal{I}_{1\setminus I}\) that doesn't increase rank, meaning \(r(A) = r(I)\). If \(r(A) = |I \setminus A| = r(I \setminus A)\), as in Eqn. (17.39), then \(A\) is in \(I\)'s span, so should get \(\text{sat}(1_I) = \text{span}(I)\).
Consider matroid \((E, \mathcal{I}) = (E, r)\), some \(I \in \mathcal{I}\). Then \(1_I \in P_r\) and

\[
\mathcal{D}(1_I) = \{ A : 1_I(A) = r(A) \} \tag{17.36}
\]

and

\[
\text{sat}(1_I) = \bigcup \{ A : A \subseteq E, A \in \mathcal{D}(1_I) \} \tag{17.37}
\]

\[
= \bigcup \{ A : A \subseteq E, 1_I(A) = r(A) \} \tag{17.38}
\]
The \textit{sat} function $=$ Polymatroid Closure

Consider matroid $(E, \mathcal{I}) = (E, r)$, some $I \in \mathcal{I}$. Then $1_I \in P_r$ and

$$D(1_I) = \{ A : 1_I(A) = r(A) \}$$

(17.36)

and

$$\text{sat}(1_I) = \bigcup \{ A : A \subseteq E, A \in D(1_I) \}$$

(17.37)

$$= \bigcup \{ A : A \subseteq E, 1_I(A) = r(A) \}$$

(17.38)

$$= \bigcup \{ A : A \subseteq E, |I \cap A| = r(A) \}$$

(17.39)
The \textit{sat} function $=$ Polymatroid Closure

- Consider matroid $(E, \mathcal{I}) = (E, r)$, some $I \in \mathcal{I}$. Then $1_I \in P_r$ and

$$\mathcal{D}(1_I) = \{ A : 1_I(A) = r(A) \} \quad (17.36)$$

and

$$\text{sat}(1_I) = \bigcup \{ A : A \subseteq E, A \in \mathcal{D}(1_I) \} \quad (17.37)$$

$$= \bigcup \{ A : A \subseteq E, 1_I(A) = r(A) \} \quad (17.38)$$

$$= \bigcup \{ A : A \subseteq E, |I \cap A| = r(A) \} \quad (17.39)$$

- Notice that $1_I(A) = |I \cap A| \leq |I|$. 
The sat function = Polymatroid Closure

- Consider matroid $(E, \mathcal{I}) = (E, r)$, some $I \in \mathcal{I}$. Then $1_I \in P_r$ and

$$\mathcal{D}(1_I) = \{ A : 1_I(A) = r(A) \}$$

(17.36)

and

$$\text{sat}(1_I) = \bigcup \{ A : A \subseteq E, A \in \mathcal{D}(1_I) \}$$

(17.37)

$$= \bigcup \{ A : A \subseteq E, 1_I(A) = r(A) \}$$

(17.38)

$$= \bigcup \{ A : A \subseteq E, |I \cap A| = r(A) \}$$

(17.39)

- Notice that $1_I(A) = |I \cap A| \leq |I|$.

- Intuitively, consider an $A \supset I \in \mathcal{I}$ that doesn’t increase rank, meaning $r(A) = r(I)$. If $r(A) = |I \cap A| = r(I \cap A)$, as in Eqn. (17.39), then $A$ is in $I$’s span, so should get $\text{sat}(1_I) = \text{span}(I)$.

\[
\cap ( (A \cup I) \cap (A \cap I) ) = r(I \cap A)
\]
The sat function = Polymatroid Closure

Consider matroid \((E, \mathcal{I}) = (E, r)\), some \(I \in \mathcal{I}\). Then \(1_I \in P_r\) and

\[
\mathcal{D}(1_I) = \{ A : 1_I(A) = r(A) \} \tag{17.36}
\]

and

\[
\text{sat}(1_I) = \bigcup \{ A : A \subseteq E, A \in \mathcal{D}(1_I) \} \tag{17.37}
\]

\[
= \bigcup \{ A : A \subseteq E, 1_I(A) = r(A) \} \tag{17.38}
\]

\[
= \bigcup \{ A : A \subseteq E, |I \cap A| = r(A) \} \tag{17.39}
\]

Notice that \(1_I(A) = |I \cap A| \leq |I|\).

Intuitively, consider an \(A \supset I \in \mathcal{I}\) that doesn’t increase rank, meaning \(r(A) = r(I)\). If \(r(A) = |I \cap A| = r(I \cap A)\), as in Eqn. (17.39), then \(A\) is in \(I\)’s span, so should get \(\text{sat}(1_I) = \text{span}(I)\).

We formalize this next.
Lemma 17.8.2 (Matroid sat : $\mathbb{R}_+^E \rightarrow 2^E$ is the same as closure.)

For $I \in \mathcal{I}$, we have $\text{sat}(1_I) = \text{span}(I)$  \hspace{1cm} (17.40)
Lemma 17.8.2 (Matroid \( \text{sat} : \mathbb{R}_+^E \rightarrow 2^E \) is the same as closure.)

For \( I \in \mathcal{I} \), we have \( \text{sat}(1_I) = \text{span}(I) \) \hspace{1cm} (17.40)

Proof.

- For \( 1_I(I) = |I| = r(I) \), so \( I \in \mathcal{D}(1_I) \) and \( I \subseteq \text{sat}(1_I) \). Also, \( I \subseteq \text{span}(I) \).
Lemma 17.8.2 (Matroid sat : $\mathbb{R}_+^E \rightarrow 2^E$ is the same as closure.)

For $I \in \mathcal{I}$, we have $\text{sat}(1_I) = \text{span}(I)$ \hfill (17.40)

Proof.

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

...
Lemma 17.8.2 (Matroid sat : \( \mathbb{R}_+^E \rightarrow 2^E \) is the same as closure.)

For \( I \in \mathcal{I} \), we have \( \text{sat}(1_I) = \text{span}(I) \) \hspace{1cm} (17.40)

Proof.

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

...
The sat function = Polymatroid Closure

Lemma 17.8.2 (Matroid sat : \( \mathbb{R}^E_+ \rightarrow 2^E \) is the same as closure.)

For \( I \in \mathcal{I} \), we have \( \text{sat}(1_I) = \text{span}(I) \) \hspace{1cm} (17.40)

Proof.

- For \( 1_I(I) = |I| = r(I) \), so \( I \in \mathcal{D}(1_I) \) and \( I \subseteq \text{sat}(1_I) \). Also, \( I \subseteq \text{span}(I) \).
- Consider some \( b \in \text{span}(I) \setminus I \).
- Then \( I \cup \{b\} \in \mathcal{D}(1_I) \) since \( 1_I(I \cup \{b\}) = |I| = r(I \cup \{b\}) = r(I) \).
- Thus, \( b \in \text{sat}(1_I) \).
Lemma 17.8.2 (Matroid \( \text{sat} : \mathbb{R}_+^E \rightarrow 2^E \) is the same as closure.)

\[
\text{For } I \in \mathcal{I}, \text{ we have } \text{sat}(1_I) = \text{span}(I) \tag{17.40}
\]

Proof.

- For \( 1_I(I) = |I| = r(I) \), so \( I \in \mathcal{D}(1_I) \) and \( I \subseteq \text{sat}(1_I) \). Also, \( I \subseteq \text{span}(I) \).

- Consider some \( b \in \text{span}(I) \setminus I \).

- Then \( I \cup \{b\} \in \mathcal{D}(1_I) \) since \( 1_I(I \cup \{b\}) = |I| = r(I \cup \{b\}) = r(I) \).

- Thus, \( b \in \text{sat}(1_I) \).

- Therefore, \( \text{sat}(1_I) \supseteq \text{span}(I) \).
...proof continued.

Now, consider $b \in \text{sat}(\mathbf{1}_I) \setminus I$. 
The sat function $=\text{Polymatroid Closure}$

...proof continued.

- Now, consider $b \in \text{sat}(1_I) \setminus I$.
- Choose any $A \in D(1_I)$ with $b \in A$, thus $b \in A \setminus I$. 
The sat function = Polymatroid Closure

...proof continued.

- Now, consider \( b \in \text{sat}(1_I) \setminus I \).
- Choose any \( A \in D(1_I) \) with \( b \in A \), thus \( b \in A \setminus I \).
- Then \( 1_I(A) = |A \cap I| = r(A) = r(A \cap I) \).
The \textit{sat} function $=$ Polymatroid Closure

... proof continued.

- Now, consider $b \in \text{sat}(1_I) \setminus I$.
- Choose any $A \in \mathcal{D}(1_I)$ with $b \in A$, thus $b \in A \setminus I$.
- Then $1_I(A) = |A \cap I| = r(A) = r(A \cap I)$.
- Now $r(A) = |A \cap I| \leq |I| = r(I)$. 
The sat function = Polymatroid Closure

... proof continued.

- Now, consider $b \in \text{sat}(1_I) \setminus I$.
- Choose any $A \in \mathcal{D}(1_I)$ with $b \in A$, thus $b \in A \setminus I$.
- Then $1_I(A) = |A \cap I| = r(A) = r(A \cap I)$.
- Now $r(A) = |A \cap I| \leq |I| = r(I)$.
- Also, $r(A \cap I) = |A \cap I|$ since $A \cap I \in \mathcal{I}$. 
The \textit{sat} function = Polymatroid Closure

...proof continued.

- Now, consider $b \in \text{sat}(1_I) \setminus I$.
- Choose any $A \in \mathcal{D}(1_I)$ with $b \in A$, thus $b \in A \setminus I$.
- Then $1_I(A) = |A \cap I| = r(A) = r(A \cap I)$.
- Now $r(A) = |A \cap I| \leq |I| = r(I)$.
- Also, $r(A \cap I) = |A \cap I|$ since $A \cap I \in \mathcal{I}$.
- Hence, $r(A \cap I) = r(A) = r((A \cap I) \cup (A \setminus I))$ meaning $(A \setminus I) \subseteq \text{span}(A \cap I) \subseteq \text{span}(I)$. 
The sat function = Polymatroid Closure

... proof continued.

- Now, consider $b \in \text{sat}(1_I) \setminus I$.
- Choose any $A \in \mathcal{D}(1_I)$ with $b \in A$, thus $b \in A \setminus I$.
- Then $1_I(A) = |A \cap I| = r(A) = r(A \cap I)$.
- Now $r(A) = |A \cap I| \leq |I| = r(I)$.
- Also, $r(A \cap I) = |A \cap I|$ since $A \cap I \in \mathcal{I}$.
- Hence, $r(A \cap I) = r(A) = r((A \cap I) \cup (A \setminus I))$ meaning $(A \setminus I) \subseteq \text{span}(A \cap I) \subseteq \text{span}(I)$.
- Since $b \in A \setminus I$, we get $b \in \text{span}(I)$. 

Prof. Jeff Bilmes
The sat function $= \text{Polymatroid Closure}$

...proof continued.

- Now, consider $b \in \text{sat}(1_I) \setminus I$.
- Choose any $A \in \mathcal{D}(1_I)$ with $b \in A$, thus $b \in A \setminus I$.
- Then $1_I(A) = |A \cap I| = r(A) = r(A \cap I)$.
- Now $r(A) = |A \cap I| \leq |I| = r(I)$.
- Also, $r(A \cap I) = |A \cap I|$ since $A \cap I \in \mathcal{I}$.
- Hence, $r(A \cap I) = r(A) = r((A \cap I) \cup (A \setminus I))$ meaning $(A \setminus I) \subseteq \text{span}(A \cap I) \subseteq \text{span}(I)$.
- Since $b \in A \setminus I$, we get $b \in \text{span}(I)$.
- Thus, $\text{sat}(1_I) \subseteq \text{span}(I)$. 

$\square$
The sat function = Polymatroid Closure

...proof continued.

- Now, consider $b \in \text{sat}(1_I) \setminus I$.
- Choose any $A \in D(1_I)$ with $b \in A$, thus $b \in A \setminus I$.
- Then $1_I(A) = |A \cap I| = r(A) = r(A \cap I)$.
- Now $r(A) = |A \cap I| \leq |I| = r(I)$.
- Also, $r(A \cap I) = |A \cap I|$ since $A \cap I \in \mathcal{I}$.
- Hence, $r(A \cap I) = r(A) = r((A \cap I) \cup (A \setminus I))$ meaning $(A \setminus I) \subseteq \text{span}(A \cap I) \subseteq \text{span}(I)$.
- Since $b \in A \setminus I$, we get $b \in \text{span}(I)$.
- Thus, $\text{sat}(1_I) \subseteq \text{span}(I)$.
- Hence $\text{sat}(1_I) = \text{span}(I)$.
Now, consider a matroid \((E, r)\) and some \(C \subseteq E\) with \(C \notin \mathcal{I}\), and consider \(1_C\).
Now, consider a matroid \((E, r)\) and some \(C \subseteq E\) with \(C \notin \mathcal{I}\), and consider \(1_C\). Is \(1_C \in P_r\)?
Now, consider a matroid \((E, \mathcal{I})\) and some \(C \subseteq E\) with \(C \notin \mathcal{I}\), and consider \(1_C\). Is \(1_C \in P_r\)? No, it is not a vertex, or even a member, of \(P_r\).
Now, consider a matroid \((E, r)\) and some \(C \subseteq E\) with \(C \notin \mathcal{I}\), and consider \(1_C\). Is \(1_C \in P_r\)? No, it is not a vertex, or even a member, of \(P_r\).

\(\text{span}(\cdot)\) operates on more than just independent sets, so \(\text{span}(C)\) is perfectly sensible.
The sat function = Polymatroid Closure

- Now, consider a matroid \((E, r)\) and some \(C \subseteq E\) with \(C \notin \mathcal{I}\), and consider \(1_C\). Is \(1_C \in P_r\)? No, it is not a vertex, or even a member, of \(P_r\).
- \(\text{span}(\cdot)\) operates on more than just independent sets, so \(\text{span}(C)\) is perfectly sensible.
- Note \(\text{span}(C) = \text{span}(B)\) where \(\mathcal{I} \ni B \in \mathcal{B}(C)\) is a base of \(C\).
The sat function = Polymatroid Closure

- Now, consider a matroid \((E, r)\) and some \(C \subseteq E\) with \(C \notin \mathcal{I}\), and consider \(1_C\). Is \(1_C \in P_r\)? No, it is not a vertex, or even a member, of \(P_r\).
- \(\text{span}()\) operates on more than just independent sets, so \(\text{span}(C)\) is perfectly sensible.
- Note \(\text{span}(C) = \text{span}(B)\) where \(\mathcal{I} \ni B \in \mathcal{B}(C)\) is a base of \(C\).
- Then we have \(1_B \leq 1_C \leq 1_{\text{span}(C)}\), and that \(1_B \in P_r\). We can then make the definition:

\[
\text{sat}(1_C) \triangleq \text{sat}(1_B) \text{ for } B \in \mathcal{B}(C) \tag{17.41}
\]

In which case, we also get \(\text{sat}(1_C) = \text{span}(C)\) (in general, could define \(\text{sat}(y) = \text{sat}(\text{P-basis}(y))\)).
The sat function = Polymatroid Closure

- Now, consider a matroid \((E, r)\) and some \(C \subseteq E\) with \(C \notin \mathcal{I}\), and consider \(1_C\). Is \(1_C \in P_r\)? No, it is not a vertex, or even a member, of \(P_r\).
- \(\text{span}(\cdot)\) operates on more than just independent sets, so \(\text{span}(C)\) is perfectly sensible.
- Note \(\text{span}(C) = \text{span}(B)\) where \(\mathcal{I} \ni B \in \mathcal{B}(C)\) is a base of \(C\).
- Then we have \(1_B \leq 1_C \leq 1_{\text{span}(C)}\), and that \(1_B \in P_r\). We can then make the definition:

\[
sat(1_C) \triangleq sat(1_B) \text{ for } B \in \mathcal{B}(C) \quad (17.41)
\]

In which case, we also get \(sat(1_C) = \text{span}(C)\) (in general, could define \(sat(y) = sat(\text{P-basis}(y))\)).
- However, consider the following form

\[
sat(1_C) = \bigcup \{A : A \subseteq E, |A \cap C| = r(A)\} \quad (17.42)
\]

Exercise: is \(\text{span}(C) = sat(1_C)\)? Prove or disprove it.
The \textit{sat} function, span, and submodular function minimization

- Thus, for a matroid, \( \text{sat}(1_I) \) is exactly the closure (or span) of \( I \) in the matroid. I.e., for matroid \((E, r)\), we have \( \text{span}(I) = \text{sat}(1_B) \).
The sat function, span, and submodular function minimization

- Thus, for a matroid, \( \text{sat}(1_I) \) is exactly the closure (or span) of \( I \) in the matroid. I.e., for matroid \((E, r)\), we have \( \text{span}(I) = \text{sat}(1_B) \).

- Recall, for \( x \in P_f \) and polymatroidal \( f \), \( \text{sat}(x) \) is the maximal (by inclusion) minimizer of \( f(A) - x(A) \), and thus in a matroid, \( \text{span}(I) \) is the maximal minimizer of the submodular function formed by \( r(A) - 1_I(A) \).
The sat function, span, and submodular function minimization

- Thus, for a matroid, \( \text{sat}(1_I) \) is exactly the closure (or span) of \( I \) in the matroid. I.e., for matroid \((E, r)\), we have \( \text{span}(I) = \text{sat}(1_B) \).
- Recall, for \( x \in P_f \) and polymatroidal \( f \), \( \text{sat}(x) \) is the maximal (by inclusion) minimizer of \( f(A) - x(A) \), and thus in a matroid, \( \text{span}(I) \) 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.
We are given an \( x \in P^+_f \) for submodular function \( f \).
We are given an $x \in P_f^+$ for submodular function $f$.

Recall that for such an $x$, $sat(x)$ is defined as

$$sat(x) = \bigcup \{ A : x(A) = f(A) \}$$  \hspace{1cm} (17.43)
We are given an \( x \in P_f^+ \) for submodular function \( f \).

Recall that for such an \( x \), \( \text{sat}(x) \) is defined as

\[
\text{sat}(x) = \bigcup \{ A : x(A) = f(A) \}
\]  \hspace{1cm} (17.43)

We also have stated that \( \text{sat}(x) \) can be defined as:

\[
\text{sat}(x) = \left\{ e : \forall \alpha > 0, x + \alpha 1_e \notin P_f^+ \right\}
\]  \hspace{1cm} (17.44)
We are given an $x \in P_f^+$ for submodular function $f$.

Recall that for such an $x$, $\text{sat}(x)$ is defined as

$$\text{sat}(x) = \bigcup \{A : x(A) = f(A)\} \quad (17.43)$$

We also have stated that $\text{sat}(x)$ can be defined as:

$$\text{sat}(x) = \left\{ e : \forall \alpha > 0, x + \alpha \mathbf{1}_e \notin P_f^+ \right\} \quad (17.44)$$

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

- Lets start with one definition and derive the other.

\[ \text{sat}(x) \]
Let's start with one definition and derive the other.

\[ \text{sat}(x) \overset{\text{def}}{=} \left\{ e : \forall \alpha > 0, x + \alpha \mathbf{1}_e \notin P_f^+ \right\} \]  (17.45)
Let's start with one definition and derive the other.

\[
sat(x) \overset{\text{def}}{=} \left\{ e : \forall \alpha > 0, x + \alpha 1_e \notin P_f^+ \right\} \\
= \left\{ e : \forall \alpha > 0, \exists A \text{ s.t. } (x + \alpha 1_e)(A) > f(A) \right\}
\] (17.45) (17.46)
sat, as tight polymatroidal elements

- Lets start with one definition and derive the other.

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

(17.45)

\[
= \left\{ e : \forall \alpha > 0, \exists A \text{ s.t. } (x + \alpha \mathbf{1}_e)(A) > f(A) \right\}
\]

(17.46)

\[
= \left\{ e : \forall \alpha > 0, \exists A \ni e \text{ s.t. } (x + \alpha \mathbf{1}_e)(A) > f(A) \right\}
\]

(17.47)
Let's start with one definition and derive the other.

\[ \text{sat}(x) \overset{\text{def}}{=} \left\{ e : \forall \alpha > 0, x + \alpha \mathbf{1}_e \notin P_f^+ \right\} \]  
(17.45)

\[ = \left\{ e : \forall \alpha > 0, \exists A \text{ s.t. } (x + \alpha \mathbf{1}_e)(A) > f(A) \right\} \]  
(17.46)

\[ = \left\{ e : \forall \alpha > 0, \exists A \exists e \text{ s.t. } (x + \alpha \mathbf{1}_e)(A) > f(A) \right\} \]  
(17.47)

This last bit follows since \( \mathbf{1}_e(A) = 1 \iff e \in A. \)
Lets start with one definition and derive the other.

\[
sat(x) \overset{\text{def}}{=} \left\{ e : \forall \alpha > 0, x + \alpha 1_e \notin P_f^{+} \right\} \quad (17.45)
\]

\[
= \left\{ e : \forall \alpha > 0, \exists A \text{ s.t. } (x + \alpha 1_e)(A) > f(A) \right\} \quad (17.46)
\]

\[
= \left\{ e : \forall \alpha > 0, \exists A \ni e \text{ s.t. } (x + \alpha 1_e)(A) > f(A) \right\} \quad (17.47)
\]

This last bit follows since \(1_e(A) = 1 \iff e \in A\). Continuing, we get

\[
sat(x) = \left\{ e : \forall \alpha > 0, \exists A \ni e \text{ s.t. } x(A) + \alpha > f(A) \right\} \quad (17.48)
\]
Let's start with one definition and derive the other.

\[
\text{sat}(x) \overset{\text{def}}{=} \left\{ e : \forall \alpha > 0, x + \alpha \mathbf{1}_e \notin P^+_f \right\} 
\]

(17.45)

\[
= \left\{ e : \forall \alpha > 0, \exists A \text{ s.t. } (x + \alpha \mathbf{1}_e)(A) > f(A) \right\} 
\]

(17.46)

\[
= \left\{ e : \forall \alpha > 0, \exists A \ni e \text{ s.t. } (x + \alpha \mathbf{1}_e)(A) > f(A) \right\} 
\]

(17.47)

This last bit follows since \( \mathbf{1}_e(A) = 1 \iff e \in A \). Continuing, we get

\[
\text{sat}(x) = \left\{ e : \forall \alpha > 0, \exists A \ni e \text{ s.t. } x(A) + \alpha > f(A) \right\} 
\]

(17.48)

given that \( x \in P^+_f \), meaning \( x(A) \leq f(A) \) for all \( A \), we must have

\[
\text{sat}(x)
\]
sat, as tight polymatroidal elements

- Let's start with one definition and derive the other.

\[
sat(x) \overset{\text{def}}{=} \left\{ e : \forall \alpha > 0, x + \alpha 1_e \notin P_f^+ \right\} \quad (17.45)
\]

\[
= \left\{ e : \forall \alpha > 0, \exists A \text{ s.t. } (x + \alpha 1_e)(A) > f(A) \right\} \quad (17.46)
\]

\[
= \left\{ e : \forall \alpha > 0, \exists A \ni e \text{ s.t. } (x + \alpha 1_e)(A) > f(A) \right\} \quad (17.47)
\]

- This last bit follows since \(1_e(A) = 1 \iff e \in A\). Continuing, we get

\[
sat(x) = \left\{ e : \forall \alpha > 0, \exists A \ni e \text{ s.t. } x(A) + \alpha > f(A) \right\} \quad (17.48)
\]

- Given that \(x \in P_f^+\), meaning \(x(A) \leq f(A)\) for all \(A\), we must have

\[
sat(x) = \left\{ e : \forall \alpha > 0, \exists A \ni e \text{ s.t. } x(A) = f(A) \right\} \quad (17.49)
\]
Let's start with one definition and derive the other.

\[
\text{sat}(x) \overset{\text{def}}{=} \left\{ e : \forall \alpha > 0, x + \alpha 1_e \notin P_f^+ \right\} \tag{17.45}
\]

\[
= \left\{ e : \forall \alpha > 0, \exists A \text{ s.t. } (x + \alpha 1_e)(A) > f(A) \right\} \tag{17.46}
\]

\[
= \left\{ e : \forall \alpha > 0, \exists A \ni e \text{ s.t. } (x + \alpha 1_e)(A) > f(A) \right\} \tag{17.47}
\]

This last bit follows since \(1_e(A) = 1 \iff e \in A\). Continuing, we get

\[
\text{sat}(x) = \left\{ e : \forall \alpha > 0, \exists A \ni e \text{ s.t. } x(A) + \alpha > f(A) \right\} \tag{17.48}
\]

given that \(x \in P_f^+\), meaning \(x(A) \leq f(A)\) for all \(A\), we must have

\[
\text{sat}(x) = \left\{ e : \forall \alpha > 0, \exists A \ni e \text{ s.t. } x(A) = f(A) \right\} \tag{17.49}
\]

\[
= \left\{ e : \exists A \ni e \text{ s.t. } x(A) = f(A) \right\} \tag{17.50}
\]
Let's start with one definition and derive the other.

\[
sat(x) \overset{\text{def}}{=} \left\{ e : \forall \alpha > 0, x + \alpha 1_e \notin P_f^+ \right\} \tag{17.45}
\]

\[
= \left\{ e : \forall \alpha > 0, \exists A \text{ s.t. } (x + \alpha 1_e)(A) > f(A) \right\} \tag{17.46}
\]

\[
= \left\{ e : \forall \alpha > 0, \exists A \ni e \text{ s.t. } (x + \alpha 1_e)(A) > f(A) \right\} \tag{17.47}
\]

This last bit follows since \(1_e(A) = 1 \iff e \in A\). Continuing, we get

\[
sat(x) = \left\{ e : \forall \alpha > 0, \exists A \ni e \text{ s.t. } x(A) + \alpha > f(A) \right\} \tag{17.48}
\]

given that \(x \in P_f^+,\) meaning \(x(A) \leq f(A)\) for all \(A\), we must have

\[
sat(x) = \left\{ e : \forall \alpha > 0, \exists A \ni e \text{ s.t. } x(A) = f(A) \right\} \tag{17.49}
\]

\[
= \left\{ e : \exists A \ni e \text{ s.t. } x(A) = f(A) \right\} \tag{17.50}
\]

So now, if \(A\) is any set such that \(x(A) = f(A)\), then we clearly have

\[
(17.51)
\]
sat, as tight polymatroidal elements

- Let's start with one definition and derive the other.

\[
sat(x) \overset{\text{def}}{=} \left\{ e : \forall \alpha > 0, x + \alpha 1_e \notin P_f^+ \right\} \tag{17.45}
\]

\[
= \left\{ e : \forall \alpha > 0, \exists A \text{ s.t. } (x + \alpha 1_e)(A) > f(A) \right\} \tag{17.46}
\]

\[
= \left\{ e : \forall \alpha > 0, \exists A \ni e \text{ s.t. } (x + \alpha 1_e)(A) > f(A) \right\} \tag{17.47}
\]

- This last bit follows since \( 1_e(A) = 1 \iff e \in A \). Continuing, we get

\[
sat(x) = \left\{ e : \forall \alpha > 0, \exists A \ni e \text{ s.t. } x(A) + \alpha > f(A) \right\} \tag{17.48}
\]

- Given that \( x \in P_f^+ \), meaning \( x(A) \leq f(A) \) for all \( A \), we must have

\[
sat(x) = \left\{ e : \forall \alpha > 0, \exists A \ni e \text{ s.t. } x(A) = f(A) \right\} \tag{17.49}
\]

\[
= \left\{ e : \exists A \ni e \text{ s.t. } x(A) = f(A) \right\} \tag{17.50}
\]

- So now, if \( A \) is any set such that \( x(A) = f(A) \), then we clearly have

\[
\forall e \in A, e \in sat(x), \tag{17.51}
\]
sat, as tight polymatroidal elements

- Lets start with one definition and derive the other.

\[
sat(x) \overset{\text{def}}{=} \left\{ e : \forall \alpha > 0, x + \alpha 1_e \notin P_f^+ \right\} \tag{17.45}
\]

\[
= \left\{ e : \forall \alpha > 0, \exists A \text{ s.t. } (x + \alpha 1_e)(A) > f(A) \right\} \tag{17.46}
\]

\[
= \left\{ e : \forall \alpha > 0, \exists A \ni e \text{ s.t. } (x + \alpha 1_e)(A) > f(A) \right\} \tag{17.47}
\]

- this last bit follows since \(1_e(A) = 1 \iff e \in A\). Continuing, we get

\[
sat(x) = \left\{ e : \forall \alpha > 0, \exists A \ni e \text{ s.t. } x(A) + \alpha > f(A) \right\} \tag{17.48}
\]

- given that \(x \in P_f^+\), meaning \(x(A) \leq f(A)\) for all \(A\), we must have

\[
sat(x) = \left\{ e : \forall \alpha > 0, \exists A \ni e \text{ s.t. } x(A) = f(A) \right\} \tag{17.49}
\]

\[
= \left\{ e : \exists A \ni e \text{ s.t. } x(A) = f(A) \right\} \tag{17.50}
\]

- So now, if \(A\) is any set such that \(x(A) = f(A)\), then we clearly have

\[
\forall e \in A, e \in sat(x), \text{ and therefore that } sat(x) \supseteq A \tag{17.51}
\]
...and therefore, with $sat$ as defined in Eq. (??),

$$sat(x) \supseteq \bigcup \{A : x(A) = f(A)\}$$  \hspace{1cm} (17.52)
...and therefore, with \( \text{sat} \) as defined in Eq. (??),

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

(17.52)

On the other hand, for any \( e \in \text{sat}(x) \) defined as in Eq. (17.50), since \( e \) is itself a member of a tight set, there is a set \( A \ni e \) such that \( x(A) = f(A) \), giving

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

(17.53)
sat, as tight polymatroidal elements

...and therefore, with sat as defined in Eq. (??),

\[
\text{sat}(x) \supseteq \bigcup \{ A : x(A) = f(A) \} \tag{17.52}
\]

On the other hand, for any \( e \in \text{sat}(x) \) defined as in Eq. (17.50), since \( e \) is itself a member of a tight set, there is a set \( A \ni e \) such that \( x(A) = f(A) \), giving

\[
\text{sat}(x) \subseteq \bigcup \{ A : x(A) = f(A) \} \tag{17.53}
\]

Therefore, the two definitions of sat are identical.
Another useful concept is saturation capacity which we develop next.
Another useful concept is saturation capacity which we develop next.

For $x \in P_f$, and $e \in E$, consider finding

$$\max \{\alpha : \alpha \in \mathbb{R}, x + \alpha 1_e \in P_f\}$$ (17.54)
Another useful concept is saturation capacity which we develop next.

For \( x \in P_f \), and \( e \in E \), consider finding

\[
\max \{ \alpha : \alpha \in \mathbb{R}, x + \alpha 1_e \in P_f \} \quad (17.54)
\]

This is identical to:

\[
\max \{ \alpha : (x + \alpha 1_e)(A) \leq f(A), \forall A \supseteq \{e\} \} \quad (17.55)
\]

since any \( B \subseteq E \) such that \( e \notin B \) does not change in a \( 1_e \) adjustment, meaning \((x + \alpha 1_e)(B) = x(B)\).
Another useful concept is saturation capacity which we develop next.

For $x \in P_f$, and $e \in E$, consider finding

$$\max \{ \alpha : \alpha \in \mathbb{R}, x + \alpha \mathbf{1}_e \in P_f \} \quad (17.54)$$

This is identical to:

$$\max \{ \alpha : (x + \alpha \mathbf{1}_e)(A) \leq f(A), \forall A \supseteq \{e\} \} \quad (17.55)$$

since any $B \subseteq E$ such that $e \notin B$ does not change in a $\mathbf{1}_e$ adjustment, meaning $(x + \alpha \mathbf{1}_e)(B) = x(B)$.

Again, this is identical to:

$$\max \{ \alpha : x(A) + \alpha \leq f(A), \forall A \supseteq \{e\} \} \quad (17.56)$$
Another useful concept is **saturation capacity** which we develop next.

For \( x \in P_f \), and \( e \in E \), consider finding

\[
\max \{ \alpha : \alpha \in \mathbb{R}, x + \alpha 1_e \in P_f \} \tag{17.54}
\]

This is identical to:

\[
\max \{ \alpha : (x + \alpha 1_e)(A) \leq f(A), \forall A \supseteq \{e\} \} \tag{17.55}
\]

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:

\[
\max \{ \alpha : x(A) + \alpha \leq f(A), \forall A \supseteq \{e\} \} \tag{17.56}
\]

or

\[
\max \{ \alpha : \alpha \leq f(A) - x(A), \forall A \supseteq \{e\} \} \tag{17.57}
\]
Saturation Capacity

- The max is achieved when

\[ \alpha = \hat{c}(x; e) \overset{\text{def}}{=} \min \{ f(A) - x(A), \forall A \supseteq \{e\} \} \]  

(17.58)
Saturation Capacity

- The max is achieved when

\[ \alpha = \hat{c}(x; e) \overset{\text{def}}{=} \min \{ f(A) - x(A), \forall A \supseteq \{e\} \} \]  

(17.58)

- \( \hat{c}(x; e) \) is known as the saturation capacity associated with \( x \in P_f \) and \( e \).
Saturation Capacity

- The max is achieved when
  \[ \alpha = \hat{c}(x; e) \overset{\text{def}}{=} \min \{ f(A) - x(A), \forall A \supseteq \{e\} \} \]  
  \hspace{1cm} (17.58)

- \( \hat{c}(x; e) \) is known as the saturation capacity associated with \( x \in P_f \) and \( e \).

- Thus we have for \( x \in P_f \),
  \[ \hat{c}(x; e) \overset{\text{def}}{=} \min \{ f(A) - x(A), \forall A \ni e \} \]  
  \hspace{1cm} (17.59)

  \[ = \max \{ \alpha : \alpha \in \mathbb{R}, x + \alpha \mathbf{1}_e \in P_f \} \]  
  \hspace{1cm} (17.60)

- More on Matroids

- Submodular Max and polyhedral approaches

- Most Violated

- Multilinear Extension

- Lovász extension examples
Saturation Capacity

- The max is achieved when
  \[ \alpha = \hat{c}(x; e) \overset{\text{def}}{=} \min \{ f(A) - x(A), \forall A \supseteq \{ e \} \} \quad (17.58) \]

- \( \hat{c}(x; e) \) is known as the saturation capacity associated with \( x \in P_f \) and \( e \).

- Thus we have for \( x \in P_f \),

  \[ \hat{c}(x; e) \overset{\text{def}}{=} \min \{ f(A) - x(A), \forall A \ni e \} \]
  \[ = \max \{ \alpha : \alpha \in \mathbb{R}, x + \alpha \mathbf{1}_e \in P_f \} \quad (17.59) \]

- We immediately see that for \( e \in E \setminus \text{sat}(x) \), we have that \( \hat{c}(x; e) > 0 \).
Saturation Capacity

- The max is achieved when
  \[ \alpha = \hat{c}(x; e) \overset{\text{def}}{=} \min \{ f(A) - x(A), \forall A \supseteq \{e\} \} \]  
  (17.58)

- \( \hat{c}(x; e) \) is known as the saturation capacity associated with \( x \in P_f \) and \( e \).

- Thus we have for \( x \in P_f \),
  \[ \hat{c}(x; e) \overset{\text{def}}{=} \min \{ f(A) - x(A), \forall A \supseteq e \} \]  
  (17.59)

  \[ = \max \{ \alpha : \alpha \in \mathbb{R}, x + \alpha \mathbf{1}_e \in P_f \} \]  
  (17.60)

- We immediately see that for \( e \in E \setminus \text{sat}(x) \), we have that \( \hat{c}(x; e) > 0 \).
- Also, we have that: \( e \in \text{sat}(x) \iff \hat{c}(x; e) = 0 \).
Saturation Capacity

- The max is achieved when

\[ \alpha = \hat{c}(x; e) \overset{\text{def}}{=} \min \{ f(A) - x(A), \forall A \supseteq \{e\} \} \quad (17.58) \]

- \( \hat{c}(x; e) \) is known as the saturation capacity associated with \( x \in P_f \) and \( e \).

- Thus we have for \( x \in P_f \),

\[ \hat{c}(x; e) \overset{\text{def}}{=} \min \{ f(A) - x(A), \forall A \supseteq \{e\} \} \]
\[ = \max \{ \alpha : \alpha \in \mathbb{R}, x + \alpha 1_e \in P_f \} \quad (17.59) \]

- 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 \( x + \alpha 1_e \in P_f \).
Saturation Capacity

• The max is achieved when

\[ \alpha = \hat{c}(x; e) \overset{\text{def}}{=} \min \{ f(A) - x(A), \forall A \supseteq \{e\} \} \]  

(17.58)

• \( \hat{c}(x; e) \) is known as the saturation capacity associated with \( x \in P_f \) and \( e \).

• Thus we have for \( x \in P_f \),

\[ \hat{c}(x; e) \overset{\text{def}}{=} \min \{ f(A) - x(A), \forall A \ni e \} \]

\[ = \max \{ \alpha : \alpha \in \mathbb{R}, x + \alpha \mathbf{1}_e \in P_f \} \]  

(17.59)

(17.60)

• We immediately see that for \( e \in E \setminus \text{sat}(x) \), we have that \( \hat{c}(x; e) > 0 \).

• Also, we have that: \( e \in \text{sat}(x) \Leftrightarrow \hat{c}(x; e) = 0 \).

• Note that any \( \alpha \) with \( 0 \leq \alpha \leq \hat{c}(x; e) \) we have \( x + \alpha \mathbf{1}_e \in P_f \).

• We also see that computing \( \hat{c}(x; e) \) is a form of submodular function minimization.