Submodular Functions, Optimization, and Applications to Machine Learning  
— Fall Quarter, Lecture 18 —  
http://www.ee.washington.edu/people/faculty/bilmes/classes/ee563_spring_2018/

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

Dec 2nd, 2020

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

Clockwise from top left: László Lovász
Jack Edmonds
Satoru Fujishige
George Nemhauser
Laurence Wolsey
András Frank
Lloyd Shapley
H. Narayanan
Robert Bixby
William Cunningham
William Tutte
Richard Rado
Alexander Schrijver
Garrett Birkhoff
Hassler Whitney
Richard Dedekind

Class Road Map - EE563

- L1(9/30): Motivation, Applications, Definitions, Properties
- L2(10/5): Sums concave(modular), uses (diversity/costs, feature selection), information theory
- L3(10/7): Monge, More Definitions, Graph and Combinatorial Examples,
- L4(10/12): Graph & Combinatorial Examples, Matrix Rank, Properties, OtherDefs, Independence
- L5(10/14): Properties, Defs of Submodularity, Independence
- L6(10/19): Matroids, Matroid Examples, Matroid Rank,
- L8(10/26): Transversal Matroid, Matroid and representation, Dual Matroid
- L9(10/28): Other Matroid Properties, Combinatorial Geometries, Matroid and Greedy, Polyhedra, Matroid Polytopes
- L10(11/2): Matroid Polytopes, Matroids → Polymatroids
- L11(11/4): Matroids → Polymatroids, Polymatroids
- L12(11/9): Polymatroids, Polymatroids and Greedy
- L13(11/16): Polymatroids and Greedy, Possible Polytopes, Extreme Points, Cardinality Constrained Maximization
- L14(11/18): Cardinality Constrained Maximization, Curvature
- L15(11/23): Curvature, Submodular Max w. Other Constraints, Start Cont. Extensions
- L16(11/25): Submodular Max w. Other Constraints, Cont. Extensions, Lovász extension
- L17(11/30): Choquet Integration, Non-linear Measure/Aggregation, Definitions/Properties, Examples.
- L18(12/2): Multilinear Extension, Submodular Max/polyhedral, Most Violated Ineq., Matroids Closure/Sat
- L19(12/9):
- L20(12/9):
- L21(12/14): final meeting (presentations) maximization.

Last day of instruction, Fri. Dec 11th. Finals Week: Dec 12-18, 2020
Rest of class

- Homework 4 posted, due Thursday Dec 17th, 2020, 11:55pm.
- Final project paper proposal, due Sunday Dec 6th, 11:59pm.
- Final project 4-page paper and presentation slides, due Sunday Dec 13th, 11:59pm.
- Final project presentation, Monday Dec 14th, starting at 10:30am.

Final project: Read and present a recent (past 5 years) paper on submodular/supermodular optimization. Paper should have both a theoretical and practical component. What is due: (1) 4-page paper summary, and (2) 10 minute presentation about the paper, will be giving presentations on Monday 12/14/2020. You must choose your paper before the 14th (this will be HW5), and you must turn in your slides and 4-page paper (this will be HW6).

Recall, grades will be based on a combination of a final project (40%) and the four homeworks (60%).

One slide review of convex closure/L.E./CI

- Convex closure: \( \hat{f}(x) = \min_{p \in \Delta^n(x)} E_{S \sim p}[f(S)] \), where where \( \Delta^n(x) = \{ p \in \mathbb{R}^n : \sum_{S \subseteq V} p_S = 1, p_S \geq 0 \forall S \subseteq V, \text{ and } \sum_{S \subseteq V} p_S 1_S = x \} \)
- “Edmonds” extension: \( \hat{f}(w) = \max(wx : x \in B_f) \)
- Lovász extension: \( f_{LE}(w) = \sum_{i=1}^m \lambda_i f(E_i) \), with \( \lambda_i \) such that \( w = \sum_{i=1}^m \lambda_i 1_{E_i} \)
- \( \hat{f}(w) = \max_{\sigma \in \Pi[m]} w^\top c_\sigma \), \( \Pi[m] \) set of \( m! \) permutations of \( [m] \), \( \sigma \in \Pi[m] \) a permutation, \( c_\sigma \) vector with \( c_\sigma = f(E_{\sigma_i}) - f(E_{\sigma_{i-1}}) \), \( E_{\sigma_i} = \{e_{\sigma_1}, e_{\sigma_2}, \ldots, e_{\sigma_i}\} \).
- Choquet integral: \( C_f(w) = \sum_{i=1}^m (w_{e_i} - w_{e_{i+1}}) f(E_i) \)
- \( \tilde{f}(w) = \int_{-\infty}^{+\infty} \hat{f}(\alpha) d\alpha, \hat{f}(\alpha) = \begin{cases} f(\{w \geq \alpha\}) & \text{if } \alpha \geq 0 \\ f(\{w \geq \alpha\}) - f(E) & \text{if } \alpha < 0 \end{cases} \)
- All the same when \( f \) is submodular.
### Concave closure

- The **concave** closure is defined as:

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

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

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

- 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).

### Multilinear extension

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

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

- 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., \( \tilde{f}(x_1, x_2, \ldots, x_k, \ldots, x_n) = \alpha \tilde{f}(x_1, x_2, \ldots, x_k, \ldots, x_n) + \beta \tilde{f}(x_1, x_2, \ldots, x'_k, \ldots, x_n) \))

- This is unfortunately not concave. However there are some useful properties.
Lemma 18.3.1

Let $\tilde{f}(x)$ 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., $\frac{\partial^2 \tilde{f}}{\partial x_i \partial x_j} \leq 0$ for all $i, j \in V$ 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).

... proof continued.

- Since $\tilde{f}$ is multilinear, derivative is a simple difference when only one argument varies, i.e.,

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

$$\geq 0$$

where the final part follows due to monotonicity of each argument, i.e., $f(S + i) \geq f(S - i)$ for any $S$ and $i \in V$.

Side note: this means $\nabla \tilde{f}(x) = (\partial \tilde{f}/\partial x_{v_1}, \partial \tilde{f}/\partial x_{v_2}, \ldots, \partial \tilde{f}/\partial x_{v_n})^\top$ has an easy expression.
Multilinear extension

... proof continued.

- Second part of proof (antitome supergradient) also relies on simple consequence of multilinearity, namely multilinearity of the derivative as well. In this case

\[
\frac{\partial^2 \tilde{f}}{\partial x_j \partial x_i} = \frac{\partial \tilde{f}}{\partial x_j}(x_1, \ldots, x_{i-1}, 1, x_{i+1}, \ldots, x_n) - \frac{\partial \tilde{f}}{\partial x_j}(x_1, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_n) = E_{S \sim x}[f(S + i + j) - f(S + i - j)] - E_{S \sim x}[f(S - i + j) - f(S - i - j)] \leq 0
\]

(18.7) (18.8) (18.9) (18.10) (18.11)

since by submodularity, we have

\[
f(S + i - j) + f(S - i + j) \geq f(S + i + j) + f(S - i - j)
\]

(18.12) 

Corollary 18.3.2

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 \( \tilde{f}(x) \leq \tilde{f}(x + \epsilon 1_v) \) 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 \( g(\alpha) = \tilde{f}(x + \alpha(1_v - 1_u)) \) is 1-D convex in \( \alpha \) for any \( u \neq v \).
Review from lecture 12

The next two slide come from lecture 12.

Join ∨ and meet ∧ for $x, y \in \mathbb{R}^E_+$

• For $x, y \in \mathbb{R}^E_+$, define vectors $x \wedge y \in \mathbb{R}^E_+$ and $x \vee y \in \mathbb{R}^E_+$ such that, for all $e \in E$

  \[
  (x \vee y)(e) = \max(x(e), y(e)) \quad (18.1)
  \]

  \[
  (x \wedge y)(e) = \min(x(e), y(e)) \quad (18.2)
  \]

  Hence,

  \[
  x \vee y \triangleq \left( \max(x(e_1), y(e_1)), \max(x(e_2), y(e_2)), \ldots, \max(x(e_n), y(e_n)) \right)
  \]

  and similarly

  \[
  x \wedge y \triangleq \left( \min(x(e_1), y(e_1)), \min(x(e_2), y(e_2)), \ldots, \min(x(e_n), y(e_n)) \right)
  \]

• From this, we can define things like an lattices, and other constructs.
Vector rank, $\text{rank}(x)$, is submodular

- Recall that the matroid rank function $r(A) = \max(|I| : I \subseteq A : I \in \mathcal{I})$ is submodular.
- The vector rank function $\text{rank}(x) = \max(y(E) : y \leq x, y \in P)$ also satisfies a form of submodularity, namely one defined on the real lattice.

**Theorem 18.3.1 (vector rank and submodularity)**

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

$$\text{rank}(u) + \text{rank}(v) \geq \text{rank}(u \lor v) + \text{rank}(u \land v) \quad (18.1)$$

- Note what happens when $u, v \in \{0, 1\}^E \subseteq \mathbb{R}_+^E$.

**Submodularity on the real lattice**

- In general, we can define “submodularity” on the real lattice.

**Definition 18.3.3 (submodularity)**

A given function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be submodular (on the real lattice) if for all $x, y \in \mathbb{R}^n$ we have

$$f(x) + f(y) \geq f(x \lor y) + f(x \land y) \quad (18.13)$$

where $x \lor y$ (resp. $x \land y$) is the element-wise max (resp. min) of $x$ and $y$.

- When $f$ is twice differentiable, this condition is identical to the off-diagonal elements of the Hessian of $f$ being non-positive, i.e.,

$$\forall i \neq j, \forall x \in \mathbb{R}^n, \quad \frac{\partial^2 f(x)}{\partial x_i \partial x_j} \leq 0 \quad (18.14)$$

See Topkis “Supermodularity and Complementarity”, 1998. Note, non-positives along the diagonal are not required.
- This is neither the same condition as convexity nor concavity.
Recall quadratic structures from Lecture 4 (next frame).

**Lemma 18.3.1**

Let $M \in \mathbb{R}^{n \times n}$ be a symmetric matrix and $m \in \mathbb{R}^n$ be a vector. Then $f : 2^V \to \mathbb{R}$ defined as

$$f(X) = m^T 1_X + \frac{1}{2} 1_X^T M 1_X$$

(18.2)

is submodular iff the off-diagonal elements of $M$ are non-positive.

**Proof.**

- Given a complete graph $G = (V, E)$, recall that $E(X)$ is the edge set with both vertices in $X \subseteq V(G)$, and that $|E(X)|$ is supermodular.
- Non-negative modular weights $w^+ : E \to \mathbb{R}_+$, $w(E(X))$ is also supermodular, so $-w(E(X))$ is submodular.
- $f$ is a modular function $m^T 1_A = m(A)$ added to a weighted submodular function, hence $f$ is submodular.

...
Example non-convex non-concave submodular quadratic

- Simply set \( M = \begin{pmatrix} 1 & -2 \\ -2 & 1 \end{pmatrix} \).

- Then \( f(x) = x^T M x \) is neither convex nor concave but it is submodular since eigenvalues are 3 and -1.

- But \( f \) is submodular on the reals since the off-diagonals of the Hessian are negative.

- Note if we define \( f'(A) = f(1_A) \), then \( f' : 2^V \rightarrow \mathbb{R} \) is a submodular set function, hence \( f \) is a continuous extension of \( f' \).

Alternate Definition

- We can equivalently define a submodular function using a definition analogous to the four-points property.

Definition 18.3.4 (submodularity)

A given function \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) is said to be submodular (on the real lattice) if for all \( x \in \mathbb{R}^n \) and elements \( i, j \in V \) and non-negative real values \( \alpha_i, \alpha_j \geq 0 \) we have that

\[
f(x + \alpha_i 1_i) + f(x + \alpha_j 1_j) \geq f(x) + f(x + \alpha_i 1_i + \alpha_j 1_j)
\]  

Proposition 18.3.5

The two definitions of submodularity on the real lattice are the same.
Definition 18.3.6
A given function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be DR-submodular if for all $x, y \in \mathbb{R}^n$ with $x \leq y$ (element wise) and for any $i \in V$ and $\alpha > 0$, we have

$$f(x + \alpha 1_i) - f(x) \geq f(y + \alpha 1_i) - f(y)$$  \hspace{1cm} (18.16)

- For twice-differential functions,
  - When $f$ is twice differentiable, there is a similar property to the above. Namely, DR-submodularity is in this case identical to all the elements of the Hessian of $f$ being non-positive, i.e.,

$$\forall i, j, \forall x \in \mathbb{R}^n, \quad \frac{\partial^2 f(x)}{\partial x_i \partial x_j} \leq 0$$ \hspace{1cm} (18.17)

  Note, non-positives along the diagonal (and everywhere) are required.

- Again, neither the same condition as convexity nor concavity.

Example non-convex non-concave DR-Submodular quadratic
- Simply set $M = \begin{pmatrix} -1 & -2 \\ -2 & -1 \end{pmatrix}$.

- Then $f(x) = x^T M x$ is neither convex nor concave but it is submodular since eigenvalues are -3 and 1.

- Again, $f$ is submodular on the reals since the off-diagonals of the Hessian are negative.

- In general, DR-submodularity implies submodularity but not vice versa.
DR-submodularity Examples

- Proposition: in R2, \( f(x_1, x_2) = x_1 + x_2 \), \( f(x_1, x_2) = x_1 + x_2 \),
  \( f(x_1, x_2) = \min(x_1, x_2) \), \( f(x_1, x_2) = \max(x_1, x_2) \),
  \( f(x_1, x_2) = x_1 \ast x_2 \), and \( f(x_1, x_2) = \|x_1 - x_2\|^p \), for \( p \geq 1 \), are all DR-submodular.

- Proposition: multilinear extension of a submodular function is DR-submodular.

- DR-submodular, despite being neither convex nor concave, are possible to either exactly or approximately maximize (and thus are an interesting class of non-convex/non-concave functions). See Hassani, Soltanolkotabi, Karbasi, "Gradient Methods for Submodular Maximization", 2017, and references therein.

Submodular Max and polyhedral approaches

- 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 can achieve progress on this front using the “multilinear extension”.
Multilinear extension (review)

**Definition 18.4.1**

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

$$F(x) = \sum_{S \subseteq V} f(S) \prod_{i \in S} x_i \prod_{j \in V \setminus S} (1 - x_j)$$  \hfill (18.18)

- Note that $F(x) = E[f(\hat{x})]$ 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 18.4.2**

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 $i,j \in V, x \in [0,1]^V$.

**Combinatorial Submodular Max**

- We saw that monotone submodular max subject to a cardinality constraint has a $1 - 1/e$ approximation guarantee using the simple (combinatorial) greedy algorithm.
- We also saw that monotone submodular max subject to $p$ simultaneous matroid independence constraints has a $1/(1+p)$ approximation guarantee (so $1/2$ for one matroid). Can this be improved upon?
- The answer is yes.
Submodular Max and Polyhedral Approaches

- Basic idea for \( \max(f(A) : A \in \mathcal{I}) \): Given a set of constraints \( \mathcal{I} \), we form a polytope \( P_\mathcal{I} \) such that \( \{1_I : I \in \mathcal{I}\} \subseteq P_\mathcal{I} \) (e.g., \( P_\mathcal{I} = \text{conv}\{1_I : I \in \mathcal{I}\} \) where \( \mathcal{I} \) are the independent sets of a matroid).
- 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 \( x^* \). This can’t be done exactly but can be done with a \( 1 - 1/e \) guarantee for the continuous optimum.
- We then round \( x^* \) to a point on the hypercube, thus giving us a solution to the discrete problem. The rounding keeps the \( 1 - 1/e \) guarantee for the discrete solution.

Continuous Greedy Algorithm

- For any down-closed polytope \( P \ni 0 \), we define a continuous process where \( x_t \in P \) moves continuously from time \( t = 0 \) to \( t = 1 \) according to the following differential equation. We start at \( x_0 = 0 \in P \) and at the end \( x_1 \in P \) approximates \( \max_{x \in P_\mathcal{I}} F(x) \).

1. Start with \( x_0 \leftarrow 0 \in P \);
2. Utilize \( \frac{dx}{dt} = v_{\text{max}}(x) \) where \( v_{\text{max}}(x) = \arg\max_{y \in P} \langle y, \nabla \tilde{f}(x) \rangle \) from \( t = 0 \) to \( t = 1 \).

- Claim is that, in the above, \( x_1 \in P \) and \( \tilde{f}(x_1) \geq (1 - 1/e) \max_{x \in P_\mathcal{I}} F(x) \).
- In practice, line 2 is done using in iterative procedure analogous to gradient descent getting \( x_{i+1} \leftarrow x_i + \nu v_{\text{max}}(x_i) \) where \( \nu \) is rate parameter and \( i \) is the iteration number. Note the argmax can be easily solve using the same Edmond’s greedy algorithm we’ve discussed.
Other results

- In a 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, J. Vondrak showed that this scheme achieves the \( \frac{1}{e_f} (1 - e^{-cf}) \) 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 to avoid needing to do exponential number of function queries (so further approximations would apply, but we get polynomial number of queries in \( 1/\epsilon^2 \) for additional \( \epsilon \) approximation).

Some Important Papers Along These Lines

  - *Early paper that showing approach works for submodular welfare problem*
The next slide comes from lecture 11.

A polymatroid function’s polyhedron is a polymatroid.

**Theorem 18.5.1**

*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*

\[
y^x(E) = \text{rank}(x) \triangleq \max \left( y(E) : y \leq x, y \in P_f^+ \right)
= \min \left( x(A) + f(E \setminus A) : A \subseteq E \right) \tag{18.10}
\]

*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:

\[
\text{rank} \left( \frac{1}{\epsilon} 1_{E \setminus B} \right) = f(E \setminus B) = \max \left\{ y(E \setminus B) : y \in P_f^+ \right\} \tag{18.11}
\]

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 12.

Considering Theorem 18.5.1, the matroid case is now a special case, where we have that:

**Corollary 18.5.2**

*We have that:*

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

(18.21)

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

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

(18.19)

- Suppose we have any \( x \in \mathbb{R}^E_+ \) such that \( x \notin P_r^+ \).
- Hence, there must be a set of \( W \subseteq 2^V \), each member of which corresponds to a violated inequality, i.e., equations of the form \( x(A) > r_M(A) \) for \( A \in W \).
- The most violated inequality when \( x \) is considered w.r.t. \( P_r^+ \) corresponds to the set \( A \) that maximizes \( x(A) - r_M(A) \), i.e., the most violated inequality is valued as:

\[
\max \{ x(A) - r_M(A) : A \in W \} = \max \{ x(A) - r_M(A) : A \subseteq E \} \quad (18.20)
\]

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

\[
\min \{ r_M(A) + x(E \setminus A) : A \subseteq E \} \quad (18.21)
\]

Consider

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

(18.22)

- Suppose we have any \( x \in \mathbb{R}^E_+ \) such that \( x \notin P_f^+ \).
- Hence, there must be a set of \( W \subseteq 2^V \), each member of which corresponds to a violated inequality, i.e., equations of the form \( x(A) > r_M(A) \) for \( A \in W \).

\[ W = \{ \{1\} \} \quad W = \{ \{2\}, \{1, 2\} \} \quad W = \{ \{1, 2\} \} \]
The most violated inequality when $x$ is considered w.r.t. $P_f^+$ corresponds to the set $A$ that maximizes $x(A) - f(A)$, i.e., the most violated inequality is valued as:

$$\max \{x(A) - f(A) : A \in \mathcal{W}\} = \max \{x(A) - f(A) : A \subseteq E\} \quad (18.23)$$

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

$$\min \{f(A) + x(E \setminus A) : A \subseteq E\} \quad (18.24)$$

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.

### Review from Lecture 6

The following three slides are review from lecture 6.
Matroids, other definitions using matroid rank \( r : 2^V \rightarrow \mathbb{Z}_+ \)

**Definition 18.6.3 (closed/flat/subspace)**

A subset \( A \subseteq E \) is closed in (equivalently, a flat or a subspace of) a 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 18.6.4 (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 \( \text{span}(A) = A \), and the span of a set is closed.

**Definition 18.6.5 (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 \( a \in A, r(A \setminus \{a\}) = |A| - 1 \)).

Matroids by circuits

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 18.6.3 (Matroid by circuits)**

Let \( E \) be a set and \( C \) be a collection of subsets of \( E \) that satisfy the following three properties:

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

Several circuit definitions for matroids.

**Theorem 18.6.3 (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.

**Fundamental circuits in matroids**

**Lemma 18.6.1**

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 \( e \)).
Matroids: The Fundamental Circuit

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

Union of matroid bases of a set

Lemma 18.6.2

Let $B(D)$ 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:

$$\bigcup_{B \in B(D)} B = D. \quad (18.25)$$

Proof.

- Define $D' \triangleq \bigcup_{B \in B(D)} B \subseteq D$, suppose $\exists d \in D$ such that $d \notin D'$.
- Hence, $\forall B \in B(D)$ 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 $B(D)$, contradicting $d \notin D'$. \qed
The sat function = Polymatroid Closure

- 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 11 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

$$ D(x) = \{A : A \subseteq E, x(A) = f(A)\} \quad (18.26) $$

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

$$ D(x) = \{A : A \subseteq E, x(A) = f(A)\} = \{A : f(A) - x(A) = 0\} \quad (18.27) $$

$$ = \{A : f(A) - x(A) = 0\} \quad (18.28) $$

- 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 18.7.1**

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

**Proof.**

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

By submodularity, we have

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B) \tag{18.29}$$

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

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

---

The $\text{sat}$ function = Polymatroid Closure

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

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

$$= \bigcup \{ A : A \subseteq E, x(A) = f(A) \} \tag{18.31}$$

$$= \{ e : e \in E, \forall \alpha > 0, x + \alpha 1_e \notin P_f \} \tag{18.32}$$

- Hence, $\text{sat}(x)$ is the maximal (zero-valued) minimizer of the submodular function $f_x(A) \overset{\triangle}{=} f(A) - x(A)$.
- Eq. (18.32) says that $\text{sat}$ consists of elements of $E$ for point $x$ that are $P_f$ saturated (any additional positive movement, in that dimension, leaves $P_f$). We’ll revisit this in a few slides.
- First, we see how $\text{sat}$ generalizes matroid closure.
Consider matroid \((E, \mathcal{I}) = (E, r)\), some \(I \in \mathcal{I}\). Then \(1_I \in P_r\) and

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

and

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

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

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

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. (18.36), then \(A\) is in \(I\)'s span, so should get \(sat(1_I) = \text{span}(I)\).

We formalize this next.

**Lemma 18.7.2 (Matroid sat : \(\mathbb{R}_E^+ \rightarrow 2^E\) is the same as closure.)**

For \(I \in \mathcal{I}\), we have \(sat(1_I) = \text{span}(I)\) (18.37)

**Proof.**

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

Hence, \( \text{sat}(1_I) = \text{span}(I) \).
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 \( \text{span}(I) = \text{sat}(1_B) \).
- 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 (e.g., \( \min_{A \subseteq V} f(A) - x(A) \)) in a polymatroid.

\textit{sat}, as tight polymatroidal elements

- We are given an \( x \in P_f^+ \) for submodular function \( f \).
- Recall that for such an \( x \), \( \text{sat}(x) \) is defined as
  \[
  \text{sat}(x) = \bigcup \{ A : x(A) = f(A) \} \tag{18.40}
  \]
- We also have stated that \( \text{sat}(x) \) can be defined as:
  \[
  \text{sat}(x) = \left\{ e : \forall \alpha > 0, x + \alpha e \notin P_f^+ \right\} \tag{18.41}
  \]
- We next show more formally that these are the same.
\( \text{sat}(x) \) as tight polymatroidal elements

- Let's start with one definition and derive the other.
  \[
  \text{sat}(x) \overset{\text{def}}{=} \left\{ e : \forall \alpha > 0, x + \alpha 1_e \notin P_f^+ \right\} \tag{18.42}
  \]
  \[
  = \left\{ e : \forall \alpha > 0, \exists A \text{ s.t. } (x + \alpha 1_e)(A) > f(A) \right\} \tag{18.43}
  \]
  \[
  = \left\{ e : \forall \alpha > 0, \exists A \ni e \text{ s.t. } (x + \alpha 1_e)(A) > f(A) \right\} \tag{18.44}
  \]
  this last bit follows since \( 1_e(A) = 1 \iff e \in A \). Continuing, we get
  \[
  \text{sat}(x) = \left\{ e : \forall \alpha > 0, \exists A \ni e \text{ s.t. } x(A) + \alpha > f(A) \right\} \tag{18.45}
  \]

- Given that \( x \in P_f^+ \), meaning \( x(A) \leq f(A) \) for all \( A \), we must have
  \[
  \text{sat}(x) = \left\{ e : \forall \alpha > 0, \exists A \ni e \text{ s.t. } x(A) = f(A) \right\} \tag{18.46}
  \]
  \[
  = \left\{ e : \exists A \ni e \text{ s.t. } x(A) = f(A) \right\} \tag{18.47}
  \]
  So now, if \( A \) is any set such that \( x(A) = f(A) \), then we clearly have
  \[ \forall e \in A, e \in \text{sat}(x), \text{ and therefore that } \text{sat}(x) \supseteq A \tag{18.48} \]

- And therefore, with \( \text{sat} \) as defined in Eq. (18.32),
  \[
  \text{sat}(x) \supseteq \bigcup \{ A : x(A) = f(A) \} \tag{18.49}
  \]

- On the other hand, for any \( e \in \text{sat}(x) \) defined as in Eq. (18.47), since \( e \)
  is itself a member of a tight set, there is a set \( A \ni e \) such that \( x(A) = f(A) \), giving
  \[
  \text{sat}(x) \subseteq \bigcup \{ A : x(A) = f(A) \} \tag{18.50}
  \]

- Therefore, the two definitions of \( \text{sat} \) are identical.
Saturation Capacity

- Another useful concept is saturation capacity which we develop next.
- For \( x \in P_f \), and \( e \in E \), consider finding

  \[
  \max \{ \alpha : \alpha \in \mathbb{R}, x + \alpha e \in P_f \} \tag{18.51}
  \]

  This is identical to:

  \[
  \max \{ \alpha : (x + \alpha e)(A) \leq f(A), \forall A \supseteq \{e\} \} \tag{18.52}
  \]

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

- Again, this is identical to:

  \[
  \max \{ \alpha : x(A) + \alpha \leq f(A), \forall A \supseteq \{e\} \} \tag{18.53}
  \]

  or

  \[
  \max \{ \alpha : \alpha \leq f(A) - x(A), \forall A \supseteq \{e\} \} \tag{18.54}
  \]

The max is achieved when

\[
\alpha = \hat{c}(x; e) \overset{\text{def}}{=} \min \{ f(A) - x(A), \forall A \supseteq \{e\} \} \tag{18.55}
\]

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

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

  \[
  \hat{c}(x; e) \overset{\text{def}}{=} \min \{ f(A) - x(A), \forall A \ni e \} = \max \{ \alpha : \alpha \in \mathbb{R}, x + \alpha 1_e \in P_f \} \tag{18.56}
  \]

- We immediately see that for \( e \in E \setminus \text{sat}(x) \), we have that \( \hat{c}(x; e) > 0 \).
- Also, we have that: \( e \in \text{sat}(x) \iff \hat{c}(x; e) = 0 \).
- Note that any \( \alpha \) with \( 0 \leq \alpha \leq \hat{c}(x; e) \) we have \( x + \alpha 1_e \in P_f \).
- We also see that computing \( \hat{c}(x; e) \) is a form of submodular function minimization.