Submodular Functions, Optimization, and Applications to Machine Learning

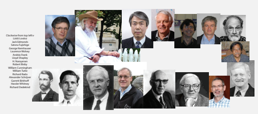
— Spring Quarter, Lecture 17 —

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

Prof. Jeff Bilmes

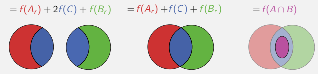
University of Washington, Seattle
Department of Electrical Engineering
http://melodi.ee.washington.edu/~bilmes

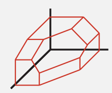
May 23st, 2018



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

$= f(A_r) + 2f(C) + f(B_r)$ $\quad = f(A_r) + f(C) + f(B_r)$ $\quad = f(A \cap B)$





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,
- L2(3/28): Machine Learning Apps (diversity, complexity, parameter, learning target, surrogate).
- 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 Submodular Defs., 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 $\rightarrow$ Polymatroids
- L10(4/29): Matroids $\rightarrow$ 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
- L15(5/14): Cont. Extensions, Lovasz Extension, Choquet Integration, Properties
- L16(5/16): More Lovasz extension, Choquet, defs/props, examples, multiliear 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):
- L21(6/4): Final Presentations maximization.

Last day of instruction, June 1st. Finals Week: June 2-8, 2018.

One slide review of concave relaxation

- convex closure $\check{f}(x) = \min_{p \in \triangle^n(x)} E_{S\sim p}[f(S)]$, where where $\triangle^n(x) = \left\{p \in \mathbb{R}^{2^n} : \sum_{S \subseteq V} p_S = 1,\ p_S \geq 0 \forall S \subseteq V,\ \&\ \sum_{S \subseteq V} p_S \mathbf{1}_S = x\right\}$
- "Edmonds" extension $\breve{f}(w) = \max(wx : x \in B_f)$
- Lovász extension $f_{\mathsf{LE}}(w) = \sum_{i=1}^m \lambda_i f(E_i)$, with λ_i such that $w = \sum_{i=1}^m \lambda_i \mathbf{1}_{E_i}$
- $\tilde{f}(w) = \max_{\sigma \in \Pi_{[m]}} w^\intercal c^\sigma$, $\Pi_{[m]}$ set of $m!$ permutations of $[m]$, $\sigma \in \Pi_{[m]}$ a permutation, c^σ vector with $c_i^\sigma = f(E_{\sigma_i}) - f(E_{\sigma_{i-1}})$, $E_{\sigma_i} = \{e_{\sigma_1}, e_{\sigma_2}, \ldots, e_{\sigma_i}\}$.
- 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.

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

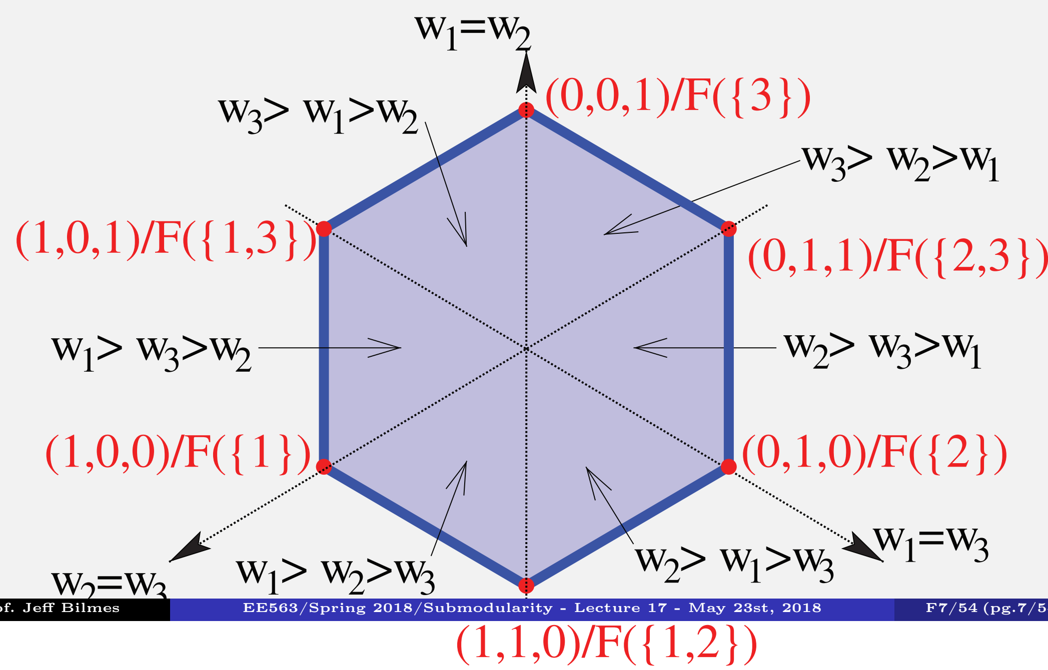
1. *Superposition of LE operator: Given f and g with Lovász extensions $\tilde{f}$ and $\tilde{g}$ then $\tilde{f} + \tilde{g}$ is the Lovász extension of $f + g$ and $\lambda \tilde{f}$ is the Lovász extension of λf for $\lambda \in \mathbb{R}$.*
2. *If $w \in \mathbb{R}_+^E$ then $\tilde{f}(w) = \int_0^{+\infty} f(\{w \geq \alpha\}) d\alpha$.*
3. *For $w \in \mathbb{R}^E$, and $\alpha \in \mathbb{R}$, we have $\tilde{f}(w + \alpha \mathbf{1}_E) = \tilde{f}(w) + \alpha f(E)$.*
4. *Positive homogeneity: I.e., $\tilde{f}(\alpha w) = \alpha \tilde{f}(w)$ for $\alpha \geq 0$.*
5. *For all $A \subseteq E$, $\tilde{f}(\mathbf{1}_A) = f(A)$.*
6. *f symmetric as in $f(A) = f(E \setminus A), \forall A$, then $\tilde{f}(w) = \tilde{f}(-w)$ ($\tilde{f}$ is even).*
7. *Given partition $E^1 \cup E^2 \cup \dots \cup E^k$ of E and $w = \sum_{i=1}^k \gamma_i \mathbf{1}_{E_k}$ with $\gamma_1 \geq \gamma_2 \geq \dots \geq \gamma_k$, and with $E^{1:i} = E^1 \cup E^2 \cup \dots \cup E^i$, 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.
- 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, $\left\{ w : \tilde{f}(w) = 1 \right\}$, 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).

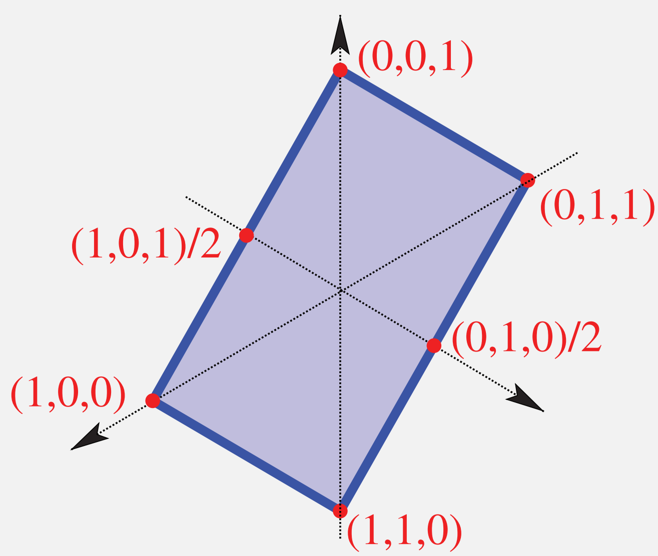
Example: $m = 3,\ E = \{1, 2, 3\}$

- 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 $\tilde{f}(w) = \max_{k \in \{1,2,3\}} w_k - \min_{k \in \{1,2,3\}} w_k$.



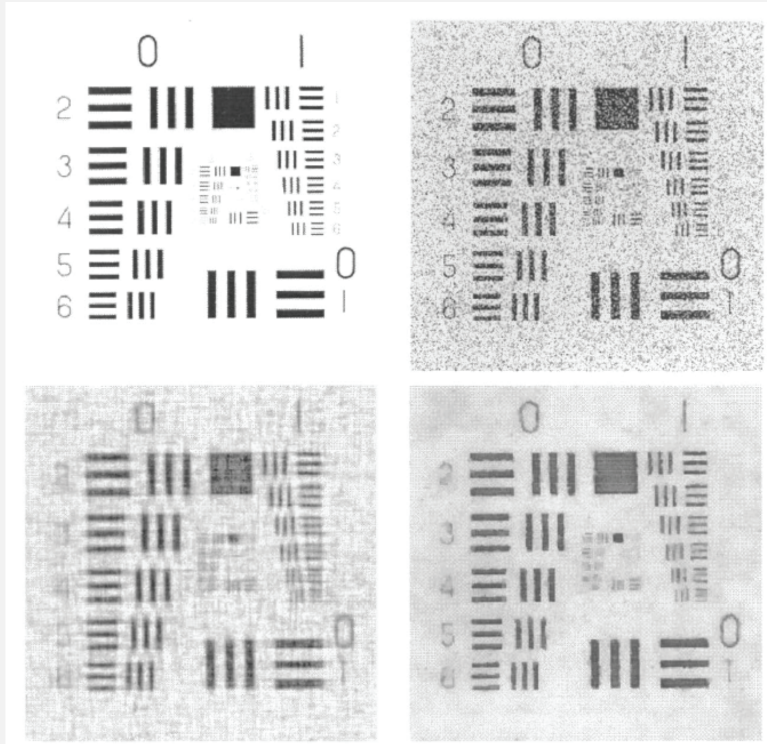
Example: $m = 3,\ E = \{1, 2, 3\}$

- Example 2 (from Bach-2011): $f(A) = |\mathbf{1}_{1\in A} - \mathbf{1}_{2 \in A}| + |\mathbf{1}_{2 \in A} - \mathbf{1}_{3 \in A}|$
- 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., $\sum_i |w_i - w_{i+1}|$).



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 $m : E \to \mathbb{R}_+$ 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

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

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

$$\tilde{f}(w) = \sum_{i=1}^{m} w(e_i)\big(g(i) - g(i-1)\big) \tag{17.2}$$

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 $m_{ij} = m((i, j))$:

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

- 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"

A few more Lovász extension examples

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 $W_k \triangleq \sum_{i=1}^{k} w(e_i)$.

$f(A)$	$\tilde{f}(w)$
$\lvert A \rvert$	$\lVert w \rVert_1$
$\min(\lvert A \rvert, 1)$	$\lVert w \rVert_\infty$
$\min(\lvert A \rvert, 1) - \max(\lvert A \rvert - m + 1, 0)$	$\lVert w \rVert_\infty - \min_i w_i$
$\min(\lvert A \rvert, k)$	W_k
$\min(\lvert A \rvert, k) - \max(\lvert A \rvert - (n - k) + 1, 1)$	$2W_k - W_m$
$\min(\lvert A \rvert, \lvert E \setminus A \rvert)$	$2W_{\lfloor m/2 \rfloor} - W_m$

(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:

$$\min_{w \in \mathbb{R}^n} \frac{1}{m} \sum_{i=1}^m \ell(y_i, w^\intercal x_i) + \lambda \Omega(w), \tag{17.5}$$

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:

$$\min_{w^1,\dots,w^k \in \mathbb{R}^n} \sum_{j=1}^k \frac{1}{m} \sum_{i=1}^m \ell(y_i^k, (w^k)^\intercal x_i) + \lambda \Omega(w^k), \tag{17.6}$$

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

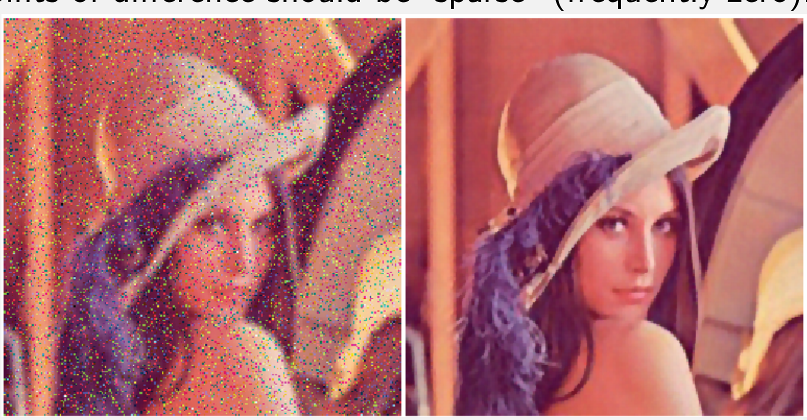
$$\min_{x_1,\dots,x_m} \min_{w^1,\dots,w^k \in \mathbb{R}^n} \sum_{j=1}^k \frac{1}{m} \sum_{i=1}^m \ell(y_i^k, (w^k)^\intercal x_i) + \lambda \Omega(w^k), \tag{17.7}$$

Norms, sparse norms, and computer vision

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

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

- 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.
- 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, $\|w\|_0 = \mathbf{1}^\intercal \text{supp}(w)$.
- 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!
- 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).
- Simple example. The Lovász extension of the modular function $f(A) = |A|$ is the ℓ_1 norm, and the Lovász extension of the modular function $f(A) = m(A)$ is the weighted ℓ_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 ℓ_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.

Concave closure

- The concave closure is defined as:

$$\hat{f}(x) = \max_{p \in \triangle^n(x)} \sum_{S \subseteq V} p_S f(S) \tag{17.9}$$

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

- This is tight at the hypercube vertices, concave, and the concave envolope 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 maxmization 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 $x \in [0,1]^V = [0,1]^{[n]}$

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

- 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 $f(\mathbf{1}_A)$ 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.

Multilinear extension

Lemma 17.4.1

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

...

Multilinear extension

... proof continued.

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

$$\frac{\partial \tilde{f}}{\partial x_v} = \tilde{f}(x_1, x_2, \dots, x_{v_1}, 1, x_{v+1}, \dots, x_n) \tag{17.11}$$

$$- \tilde{f}(x_1, x_2, \dots, x_{v_1}, 0, x_{v+1}, \dots, x_n) \tag{17.12}$$

$$= E_{S \sim x}[f(S+v)] - E_{S \sim x}[f(S-v)] \tag{17.13}$$

$$\geq 0 \tag{17.14}$$

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

$\square$

Multilinear extension

. . . proof continued.

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

$$\frac{\partial^2 \tilde{f}}{\partial x_i \partial x_j} = \frac{\partial \tilde{f}}{\partial x_j}(x_1, \dots, x_{i-1}, 1, x_{i+1}, \dots, x_n) \tag{17.15}$$

$$- \frac{\partial \tilde{f}}{\partial x_j}(x_1, \dots, x_{i-1}, 0, x_{i+1}, \dots, x_n) \tag{17.16}$$

$$= E_{S \sim x}[f(S + i + j) - f(S + i - j)] \tag{17.17}$$

$$- E_{S \sim x}[f(S - i + j) - f(S - i - j)] \tag{17.18}$$

$$\leq 0 \tag{17.19}$$

since by submodularity, we have

$$f(S + i - j) + f(S - i + j) \geq f(S + i + j) + f(S - i - j) \tag{17.20}$$

□

Multilinear extension: some properties

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 α 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 α for any $u \neq v$.*

Submodular Max and polyhedral approaches

- 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

$$F(x) = \sum_{S \subseteq V} f(S) \prod_{i \in S} x_i \prod_{j \in V \setminus S} (1 - x_j) \tag{17.21}$$

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

Submodular Max and polyhedral approaches

- Basic idea: Given a set of constraints $\mathcal{I}$, we form a polytope $P_{\mathcal{I}}$ such that $\{\mathbf{1}_I : I \in \mathcal{I}\} \subseteq P_{\mathcal{I}}$
- 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.

Submodular Max and polyhedral approaches

- 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 ℓ 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).

Review from lecture 10

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

$$y^x(E) = \textit{rank}(x) \triangleq \max\left(y(E) : y \leq x, y \in P_f^+\right)$$
$$= \min\left(x(A) + f(E \setminus A) : A \subseteq E\right) \tag{17.10}$$

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:

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

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

Review from lecture 11

The next slide comes from lecture 11.

Matroid instance of Theorem ??

- Considering Theorem **??**, the matroid case is now a special case, where we have that:

Corollary 17.6.2

We have that:

$$\max\{y(E) : y \in P_{\textit{ind. set}}(M), y \le x\} = \min\{r_M(A) + x(E \setminus A) : A \subseteq E\} \tag{17.21}$$

where r_M is the matroid rank function of some matroid.

Most violated inequality problem in matroid polytope case

- Consider

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

- 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:

$$\max\left\{x(A) - r_M(A) : A \in \mathcal{W}\right\} = \max\left\{x(A) - r_M(A) : A \subseteq E\right\} \quad (17.23)$$

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

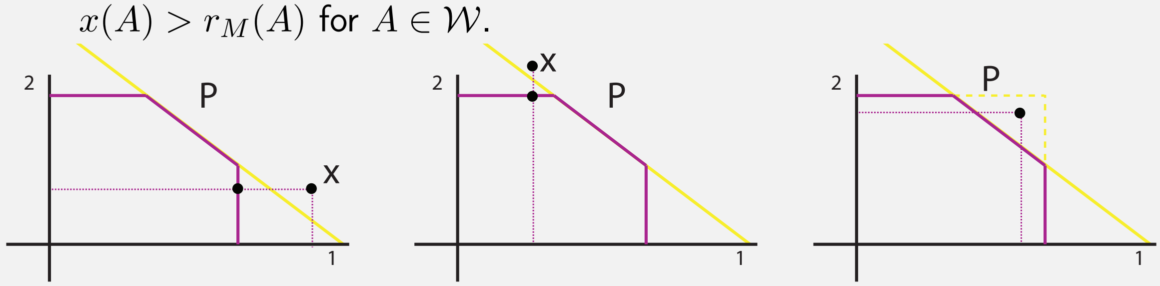
$$\min\left\{r_M(A) + x(E \setminus A) : A \subseteq E\right\} \quad (17.24)$$

Most violated inequality/polymatroid membership/SFM

- Consider

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

- 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 $A \in \mathcal{W}$.



$\mathcal{W} = \{\{1\}\{1,2\}\}$ $\quad$ $\mathcal{W} = \{\{2\},\{1,2\}\}$ $\quad$ $\mathcal{W} = \{\{1,2\}\}$

Most violated inequality/polymatroid membership/SFM

- 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).
- We will ultimatley 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 $r : 2^V \to \mathbb{Z}_+$

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

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

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

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 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 $\mathcal{C}$ be a collection of nonempty subsets of E, such that no two sets in $\mathcal{C}$ are contained in each other. Then the following are equivalent.

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

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

Fundamental circuits in matroids

Lemma 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).

Matroids: The Fundamental Circuit

- Define $C(I,e)$ be the unique circuit associated with $I \cup \{e\}$ (the fundamental circuit in M w.r.t. I and e, if it exists).
- If $e \in \text{span}(I) \setminus I$, then $C(I,e)$ is well defined ($I + e$ creates one circuit).
- If $e \in I$, then $I + e = I$ doesn't create a circuit. In such cases, $C(I,e)$ is not really defined.
- In such cases, we define $C(I,e) = \{e\}$, and we will soon see why.
- If $e \notin \text{span}(I)$ (i.e., when $I + e$ is independent), then we set $C(I,e) = \emptyset$, since no circuit is created in this case.

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 $d' \neq d$.
- 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 $d \notin D'$.

The 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.
- 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 $A, B \in \mathcal{D}(x)$, we have that $A \cup B \in \mathcal{D}(x)$ and $A \cap B \in \mathcal{D}(x)$, which can constitute a join and meet.
- Recall, for a given $x \in P_f$, we have defined this tight family as

$$\mathcal{D}(x) = \{A : A \subseteq E, x(A) = f(A)\} \tag{17.29}$$

The sat function = Polymatroid Closure

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

$$\mathcal{D}(x) = \{A : A \subseteq E, x(A) = f(A)\} \tag{17.30}$$
$$= \{A : f(A) - x(A) = 0\} \tag{17.31}$$

- 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 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 $\mathcal{M} = \operatorname{argmin}_{X \subseteq E} f(X)$ 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

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B) \tag{17.32}$$

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.

The sat function = 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:

$$\mathrm{cl}(x) \stackrel{\text{def}}{=} \mathrm{sat}(x) \stackrel{\text{def}}{=} \bigcup \{A : A \in \mathcal{D}(x)\} \tag{17.33}$$
$$= \bigcup \{A : A \subseteq E, x(A) = f(A)\} \tag{17.34}$$
$$= \{e : e \in E, \forall \alpha > 0, x + \alpha \mathbf{1}_e \notin P_f\} \tag{17.35}$$

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

The sat function = Polymatroid Closure

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

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

and

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

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

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

- Notice that $\mathbf{1}_I(A) = |I \cap A| \leq |I|$.
- 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 $\text{sat}(\mathbf{1}_I) = \text{span}(I)$.
- We formalize this next.

The sat function = Polymatroid Closure

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

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

Proof.

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

...

The sat function = Polymatroid Closure

. . . proof continued.

- Now, consider $b \in \text{sat}(\mathbf{1}_I) \setminus I$.
- Choose any $A \in \mathcal{D}(\mathbf{1}_I)$ with $b \in A$, thus $b \in A \setminus I$.
- Then $\mathbf{1}_I(A) = |A \cap I| = r(A) = r(A \cap I)$.
- 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, $\text{sat}(\mathbf{1}_I) \subseteq \text{span}(I)$.
- Hence $\text{sat}(\mathbf{1}_I) = \text{span}(I)$

□

Lovász extension examples | Multilinear Extension | Submodular Max and polyhedral approaches | Most Violated $\leq$ | More on Matroids | Closure/Sat

The sat function = Polymatroid Closure

- Now, consider a matroid (E, r) and some $C \subseteq E$ with $C \notin \mathcal{I}$, and consider $\mathbf{1}_C$. Is $\mathbf{1}_C \in P_r$? 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 $\mathbf{1}_B \leq \mathbf{1}_C \leq \mathbf{1}_{\text{span}(C)}$, and that $\mathbf{1}_B \in P_r$. We can then make the definition:

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

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

- However, consider the following form

$$\text{sat}(\mathbf{1}_C) = \bigcup \{A : A \subseteq E, |A \cap C| = r(A)\} \tag{17.42}$$

Exercise: is $\text{span}(C) = \text{sat}(\mathbf{1}_C)$? Prove or disprove it.

The sat function, span, and submodular function minimization

- Thus, for a matroid, $\text{sat}(\mathbf{1}_I)$ is exactly the closure (or span) of I in the matroid. I.e., for matroid (E, r), we have $\text{span}(I) = \text{sat}(\mathbf{1}_B)$.
- 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, span(I) is the maximal minimizer of the submodular function formed by $r(A) - \mathbf{1}_I(A)$.
- Submodular function minimization can solve "span" queries in a matroid or "sat" queries in a polymatroid.

sat, as tight polymatroidal elements

- We are given an $x \in P_f^+$ for submodular function f.
- Recall that for such an x, $\text{sat}(x)$ is defined as

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

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

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

- We next show more formally that these are the same.

sat, as tight polymatroidal elements

- Lets start with one definition and derive the other.

$$\begin{aligned} \text{sat}(x) &\stackrel{\text{def}}{=} \left\{e : \forall \alpha > 0, x + \alpha \mathbf{1}_e \notin P_f^+ \right\} && (17.45)\\ &= \{e : \forall \alpha > 0, \exists A \text{ s.t. } (x + \alpha \mathbf{1}_e)(A) > f(A)\} && (17.46)\\ &= \{e : \forall \alpha > 0, \exists A \ni e \text{ s.t. } (x + \alpha \mathbf{1}_e)(A) > f(A)\} && (17.47)\end{aligned}$$

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

$$\text{sat}(x) = \{e : \forall \alpha > 0, \exists A \ni e \text{ s.t. } x(A) + \alpha > f(A)\} \qquad (17.48)$$

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

$$\begin{aligned} \text{sat}(x) &= \{e : \forall \alpha > 0, \exists A \ni e \text{ s.t. } x(A) = f(A)\} && (17.49)\\ &= \{e : \exists A \ni e \text{ s.t. } x(A) = f(A)\} && (17.50)\end{aligned}$$

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

$$\forall e \in A, e \in \text{sat}(x), \text{ and therefore that } \text{sat}(x) \supseteq A \qquad (17.51)$$

Lovász extension examples · Multilinear Extension · Submodular Max and polyhedral approaches · Most Violated $\leq$ · More on Matroids · Closure/Sat

sat, as tight polymatroidal elements

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

$$\text{sat}(x) \supseteq \bigcup \{A : x(A) = f(A)\} \qquad (17.52)$$

- 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

$$\text{sat}(x) \subseteq \bigcup \{A : x(A) = f(A)\} \qquad (17.53)$$

- Therefore, the two definitions of 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

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

- This is identical to:

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

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

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

  or

$$\max \{\alpha : \alpha \leq f(A) - x(A), \forall A \supseteq \{e\}\} \tag{17.57}$$

Saturation Capacity

- The max is achieved when

$$\alpha = \hat{c}(x; e) \stackrel{\text{def}}{=} \min \{f(A) - x(A), \forall A \supseteq \{e\}\} \tag{17.58}$$

- $\hat{c}(x; e)$ is known as the saturation capacity associated with $x \in P_f$ and e.
- Thus we have for $x \in P_f$,

$$\hat{c}(x; e) \stackrel{\text{def}}{=} \min \{f(A) - x(A), \forall A \ni e\} \tag{17.59}$$
$$= \max \{\alpha : \alpha \in \mathbb{R}, x + \alpha \mathbf{1}_e \in P_f\} \tag{17.60}$$

- 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 α 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.