Submodular Functions, Optimization, and Applications to Machine Learning

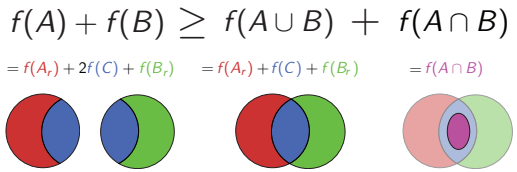
— Fall Quarter, Lecture 16 —

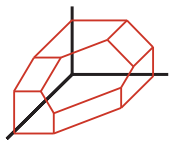
http://www.ee.washington.edu/people/faculty/bilmes/classes/ee563_spring_2018/

Prof. Jeff Bilmes

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

Nov 25th, 2020





Announcements, Assignments, and Reminders

- Homework 3, due Wednesday (tonight), Nov 25th, 2020, 11:59pm.
- Office hours this week, Monday (11/23), Tues (11/24), & Wed (11/25), 10:00pm at our class zoom link. ~~I can meet Monday night at 10:00pm as well on request.~~
- Happy Thanksgiving!

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, Other Defs, Independence
- L5(10/14): Properties, Defs of Submodularity, Independence
- L6(10/19): Matroids, Matroid Examples, Matroid Rank,
- L7(10/21): Matroid Rank, More on Partition Matroid, Laminar Matroids, System of Distinct Reps, Transversals
- 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 $\rightarrow$ Polymatroids
- L11(11/4): Matroids $\rightarrow$ Polymatroids, Polymatroids
- L12(11/9): Polymatroids, Polymatroids and Greedy
- L–(11/11): Veterans Day, Holiday
- 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):
- L18(12/2):
- L19(12/7):
- 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 will come out later this week, will be due about 1.5-2 weeks after that.
- 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).

Review

- We've been discussing algorithms and approximations for $\max_{A \in \mathcal{C}} f(A)$ where f is a polymatroid function and $\mathcal{C}$ is a constraint set (e.g., cardinality, matroid, multiple matroids, knapsacks) as well as other problems (e.g., submodular welfare problem).

$V = \bigcup_i V_i \quad V_i \subseteq V$, partition $\{V_1, V_2, \ldots, V_m\}$

Submodular Fair Allocation

SW $\max_{(V_1, \ldots, V_m)} \frac{1}{m} \sum_{i=1}^{m} f_i(V_i)$

$\max_{(V_1, V_2, \ldots, V_m)} \min_{i \in [m]} f_i(V_i)$

What About Non-monotone

- We sometimes wish to maximize non-monotone submodular functions. This includes of course graph cuts, and this problem is APX-hard, so maximizing non-monotone functions, even unconstrainedly, is hard.

Given polymatroid $f: 2^V \to \mathbb{R}$

Form non-monotone function by, e.g.,

1) $\hat{f}(A) = f(A) - m(A)$

or 2) $\bar{f}(A) = f(A) + f(V \setminus A) - f(V)$

Ex: $f(A) = H(X_A)$

$\bar{f}(A) = I(A; V \setminus A)$

What About Non-monotone

- We sometimes wish to maximize non-monotone submodular functions. This includes of course graph cuts, and this problem is APX-hard, so maximizing non-monotone functions, even unconstrainedly, is hard.
- If f is an arbitrary submodular function (so neither polymatroidal, nor necessarily positive or negative), then verifying if the maximum of f is positive or negative is already NP-hard (this is even true for the graph cut function, see Hong, "Inapproximability Of The Max-Cut Problem With Negative Weights", 2008).

$f(sol) \geq \alpha(n) \cdot OPT$ $\alpha(n) > 0$

What About Non-monotone

- We sometimes wish to maximize non-monotone submodular functions. This includes of course graph cuts, and this problem is APX-hard, so maximizing non-monotone functions, even unconstrainedly, is hard.
- If f is an arbitrary submodular function (so neither polymatroidal, nor necessarily positive or negative), then verifying if the maximum of f is positive or negative is already NP-hard (this is even true for the graph cut function, see Hong, "Inapproximability Of The Max-Cut Problem With Negative Weights", 2008).
- Therefore, submodular function max in such case is inapproximable unless P=NP (since any such procedure would give us the sign of the max).

What About Non-monotone

- We sometimes wish to maximize non-monotone submodular functions. This includes of course graph cuts, and this problem is APX-hard, so maximizing non-monotone functions, even unconstrainedly, is hard.
- If f is an arbitrary submodular function (so neither polymatroidal, nor necessarily positive or negative), then verifying if the maximum of f is positive or negative is already NP-hard (this is even true for the graph cut function, see Hong, "Inapproximability Of The Max-Cut Problem With Negative Weights", 2008).
- Therefore, submodular function max in such case is inapproximable unless P=NP (since any such procedure would give us the sign of the max).
- Thus, any approximation algorithm must be for unipolar submodular functions. E.g., non-negative but otherwise arbitrary submodular functions.

Submodularity and local optima

- Given any submodular function f, a set $S \subseteq V$ is a local maximum of f if $f(S - v) \leq f(S)$ for all $v \in S$ and $f(S + v) \leq f(S)$ for all $v \in V \setminus S$ (i.e., local in a Hamming ball of radius 1).

Submodularity and local optima

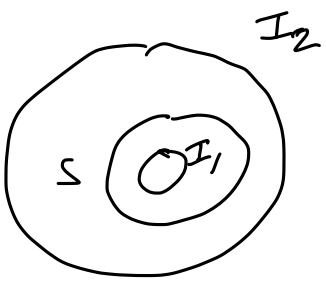
- Given any submodular function f, a set $S \subseteq V$ is a local maximum of f if $f(S - v) \leq f(S)$ for all $v \in S$ and $f(S + v) \leq f(S)$ for all $v \in V \setminus S$ (i.e., local in a Hamming ball of radius 1).
- The following interesting result is true for any submodular function:

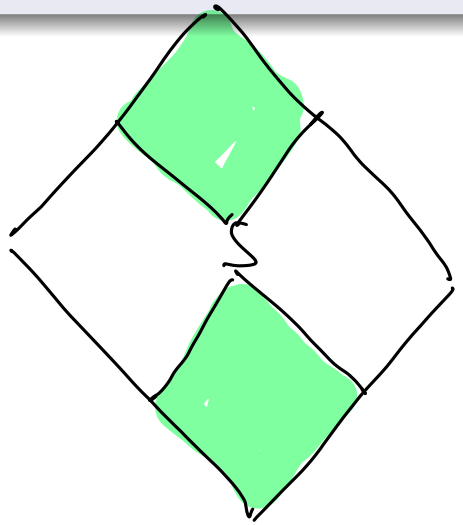
Submodularity and local optima

- Given any submodular function f, a set $S \subseteq V$ is a local maximum of f if $f(S - v) \leq f(S)$ for all $v \in S$ and $f(S + v) \leq f(S)$ for all $v \in V \setminus S$ (i.e., local in a Hamming ball of radius 1).
- The following interesting result is true for any submodular function:

Lemma 16.3.1

Given a submodular function f, if S is a local maximum of f, and $I \subseteq S$ or $I \supseteq S$, then $f(I) \leq f(S)$.





Submodularity and local optima

- Given any submodular function f, a set $S \subseteq V$ is a local maximum of f if $f(S - v) \leq f(S)$ for all $v \in S$ and $f(S + v) \leq f(S)$ for all $v \in V \setminus S$ (i.e., local in a Hamming ball of radius 1).
- The following interesting result is true for any submodular function:

Lemma 16.3.1

Given a submodular function f, if S is a local maximum of f, and $I \subseteq S$ or $I \supseteq S$, then $f(I) \leq f(S)$.

- Idea of proof: Given $v_1, v_2 \in S$, suppose $f(S - v_1) \leq f(S)$ and $f(S - v_2) \leq f(S)$. Submodularity requires $f(S - v_1) + f(S - v_2) \geq f(S) + f(S - v_1 - v_2)$ which would be impossible unless $f(S - v_1 - v_2) \leq f(S)$. I.e., if also $f(S - v_1 - v_2) > f(S)$ were true, submodularity can't hold.

Submodularity and local optima

- Given any submodular function f, a set $S \subseteq V$ is a local maximum of f if $f(S - v) \leq f(S)$ for all $v \in S$ and $f(S + v) \leq f(S)$ for all $v \in V \setminus S$ (i.e., local in a Hamming ball of radius 1).
- The following interesting result is true for any submodular function:

Lemma 16.3.1

Given a submodular function f, if S is a local maximum of f, and $I \subseteq S$ or $I \supseteq S$, then $f(I) \leq f(S)$.

- Idea of proof: Given $v_1, v_2 \in S$, suppose $f(S - v_1) \leq f(S)$ and $f(S - v_2) \leq f(S)$. Submodularity requires $f(S - v_1) + f(S - v_2) \geq f(S) + f(S - v_1 - v_2)$ which would be impossible unless $f(S - v_1 - v_2) \leq f(S)$. I.e., if also $f(S - v_1 - v_2) > f(S)$ were true, submodularity can't hold.
- Similarly, given $v_1, v_2 \notin S$, and $f(S + v_1) \leq f(S)$ and $f(S + v_2) \leq f(S)$. Submodularity requires $f(S + v_1) + f(S + v_2) \geq f(S) + f(S + v_1 + v_2)$ which requires $f(S + v_1 + v_2) \leq f(S)$.

Submodularity and local optima

- Given any submodular function f, a set $S \subseteq V$ is a local maximum of f if $f(S - v) \leq f(S)$ for all $v \in S$ and $f(S + v) \leq f(S)$ for all $v \in V \setminus S$ (i.e., local in a Hamming ball of radius 1).
- The following interesting result is true for any submodular function:

Lemma 16.3.1

Given a submodular function f, if S is a local maximum of f, and $I \subseteq S$ or $I \supseteq S$, then $f(I) \leq f(S)$.

- In other words, once we have identified a local maximum, the two intervals in the Boolean lattice $[\emptyset, S]$ and $[S, V]$ can be ruled out as a possible improvement over S.

Submodularity and local optima

- Given any submodular function f, a set $S \subseteq V$ is a local maximum of f if $f(S - v) \leq f(S)$ for all $v \in S$ and $f(S + v) \leq f(S)$ for all $v \in V \setminus S$ (i.e., local in a Hamming ball of radius 1).
- The following interesting result is true for any submodular function:

Lemma 16.3.1

Given a submodular function f, if S is a local maximum of f, and $I \subseteq S$ or $I \supseteq S$, then $f(I) \leq f(S)$.

- In other words, once we have identified a local maximum, the two intervals in the Boolean lattice $[\emptyset, S]$ and $[S, V]$ can be ruled out as a possible improvement over S.
- Finding a local maximum is already hard (PLS-complete), but it is possible to find an approximate local maximum relatively efficiently.

Submodularity and local optima

- Given any submodular function f, a set $S \subseteq V$ is a local maximum of f if $f(S - v) \leq f(S)$ for all $v \in S$ and $f(S + v) \leq f(S)$ for all $v \in V \setminus S$ (i.e., local in a Hamming ball of radius 1).
- The following interesting result is true for any submodular function:

Lemma 16.3.1

Given a submodular function f, if S is a local maximum of f, and $I \subseteq S$ or $I \supseteq S$, then $f(I) \leq f(S)$.

- In other words, once we have identified a local maximum, the two intervals in the Boolean lattice $[\emptyset, S]$ and $[S, V]$ can be ruled out as a possible improvement over S.
- Finding a local maximum is already hard (PLS-complete), but it is possible to find an approximate local maximum relatively efficiently.
- This is the approach can yield a $(\frac{1}{3} - \frac{\epsilon}{n})$ approximation algorithm for maximizing non-monotone non-negative submodular functions, with most $O(\frac{1}{\epsilon} n^3 \log n)$ function calls using approximate local maxima search.

Linear time algorithm unconstrained non-monotone max

- Tight randomized tight $1/2$ approximation algorithm for unconstrained non-monotone non-negative submodular maximization.

Linear time algorithm unconstrained non-monotone max

- Tight randomized tight $1/2$ approximation algorithm for unconstrained non-monotone non-negative submodular maximization.
- Buchbinder, Feldman, Naor, Schwartz 2012.

Linear time algorithm unconstrained non-monotone max

- Tight randomized tight $1/2$ approximation algorithm for unconstrained non-monotone non-negative submodular maximization.
- Buchbinder, Feldman, Naor, Schwartz 2012. Recall $[a]_+ = \max(a, 0)$.

Linear time algorithm unconstrained non-monotone max

- Tight randomized tight $1/2$ approximation algorithm for unconstrained non-monotone non-negative submodular maximization.
- Buchbinder, Feldman, Naor, Schwartz 2012. Recall $[a]_+ = \max(a, 0)$.

Bidirectional Greedy

Algorithm 4: Randomized Linear-time non-monotone submodular max

Set $L \leftarrow \emptyset$; $U \leftarrow V$ /* Lower L, upper U. Invariant: $L \subseteq U$ */ ;
Order elements of $V = (v_1, v_2, \ldots, v_n)$ arbitrarily ;
for $i \leftarrow 0 \ldots |V|$ **do**
  $a \leftarrow [f(v_i|L)]_+$; $b \leftarrow [-f(U|U \setminus \{v_i\})]_+$;
  if $a = b = 0$ **then** $p \leftarrow 1/2$;
  else $p \leftarrow a/(a+b)$;
  if *Flip of coin with* $\Pr(\textit{heads}) = p$ *draws heads* **then**
    $L \leftarrow L \cup \{v_i\}$;
  Otherwise /* if the coin drew tails, an event with prob. $1-p$ */
    $U \leftarrow U \setminus \{v_i\}$
return L *(which is the same as* U *at this point)*

Linear time algorithm unconstrained non-monotone max

- Each "sweep" of the algorithm is $O(n)$.

Linear time algorithm unconstrained non-monotone max

- Each "sweep" of the algorithm is $O(n)$.
- Running the algorithm $1\times$ (with an arbitrary variable order) results in a $1/3$ approximation.

Linear time algorithm unconstrained non-monotone max

- Each "sweep" of the algorithm is $O(n)$.
- Running the algorithm $1\times$ (with an arbitrary variable order) results in a $1/3$ approximation.
- The $1/2$ guarantee is in expected value (the expected solution has the $1/2$ guarantee).

Linear time algorithm unconstrained non-monotone max

- Each "sweep" of the algorithm is $O(n)$.
- Running the algorithm $1\times$ (with an arbitrary variable order) results in a $1/3$ approximation.
- The $1/2$ guarantee is in expected value (the expected solution has the $1/2$ guarantee).
- In practice, run it multiple times, each with a different random permutation of the elements, and then take the cumulative best.

Linear time algorithm unconstrained non-monotone max

- Each "sweep" of the algorithm is $O(n)$.
- Running the algorithm $1\times$ (with an arbitrary variable order) results in a $1/3$ approximation.
- The $1/2$ guarantee is in expected value (the expected solution has the $1/2$ guarantee).
- In practice, run it multiple times, each with a different random permutation of the elements, and then take the cumulative best.
- It may be possible to choose the random order smartly to get better results in practice.

Linear time algorithm unconstrained non-monotone max

- Each "sweep" of the algorithm is $O(n)$.
- Running the algorithm $1\times$ (with an arbitrary variable order) results in a $1/3$ approximation.
- The $1/2$ guarantee is in expected value (the expected solution has the $1/2$ guarantee).
- In practice, run it multiple times, each with a different random permutation of the elements, and then take the cumulative best.
- It may be possible to choose the random order smartly to get better results in practice.
- But note, even a random subset is a 1/4 approximation to the optimal solution for unconstrained non-monotone submodular maximization, in expectation (Feige, Mirrokni, Vondrak, Maximizing non-monotone submodular functions. SIAM Journal on Computing, 40(4):1133-1153, 2011.)

More general still: multiple constraints different types

- In the past several years, there has been a plethora of papers on maximizing both monotone and non-monotone submodular functions under various combinations of one or more knapsack and/or matroid constraints.

More general still: multiple constraints different types

- In the past several years, there has been a plethora of papers on maximizing both monotone and non-monotone submodular functions under various combinations of one or more knapsack and/or matroid constraints.
- The approximation quality is usually some function of the number of matroids, and is often not a function of the number of knapsacks.

More general still: multiple constraints different types

- In the past several years, there has been a plethora of papers on maximizing both monotone and non-monotone submodular functions under various combinations of one or more knapsack and/or matroid constraints.
- The approximation quality is usually some function of the number of matroids, and is often not a function of the number of knapsacks.
- Often the computational costs of the algorithms are prohibitive (e.g., exponential in k) with large constants, so these algorithms might not scale.

More general still: multiple constraints different types

- In the past several years, there has been a plethora of papers on maximizing both monotone and non-monotone submodular functions under various combinations of one or more knapsack and/or matroid constraints.
- The approximation quality is usually some function of the number of matroids, and is often not a function of the number of knapsacks.
- Often the computational costs of the algorithms are prohibitive (e.g., exponential in k) with large constants, so these algorithms might not scale.
- On the other hand, these algorithms offer deep and interesting intuition into submodular functions, beyond what we have covered here.

Some results on submodular maximization

- As we've seen, we can get $1 - 1/e$ for non-negative monotone submodular (polymatroid) functions with greedy algorithm under cardinality constraints, and this is tight.

Some results on submodular maximization

- As we've seen, we can get $1 - 1/e$ for non-negative monotone submodular (polymatroid) functions with greedy algorithm under cardinality constraints, and this is tight.
- For general matroid, greedy reduces to $1/2$ approximation (as we've seen).

Some results on submodular maximization

- As we've seen, we can get $1 - 1/e$ for non-negative monotone submodular (polymatroid) functions with greedy algorithm under cardinality constraints, and this is tight.
- For general matroid, greedy reduces to $1/2$ approximation (as we've seen).
- We can recover $1 - 1/e$ approximation using the continuous greedy algorithm on the multilinear extension and then using pipage rounding to re-integerize the solution (see J. Vondrak's publications).

Some results on submodular maximization

- As we've seen, we can get $1 - 1/e$ for non-negative monotone submodular (polymatroid) functions with greedy algorithm under cardinality constraints, and this is tight.
- For general matroid, greedy reduces to $1/2$ approximation (as we've seen).
- We can recover $1 - 1/e$ approximation using the continuous greedy algorithm on the multilinear extension and then using pipage rounding to re-integerize the solution (see J. Vondrak's publications).
- More general constraints are possible too, as we see on the next table (for references, see Jan Vondrak's publications `http://theory.stanford.edu/~jvondrak/`).

Venn Family of Subclusive Constraints

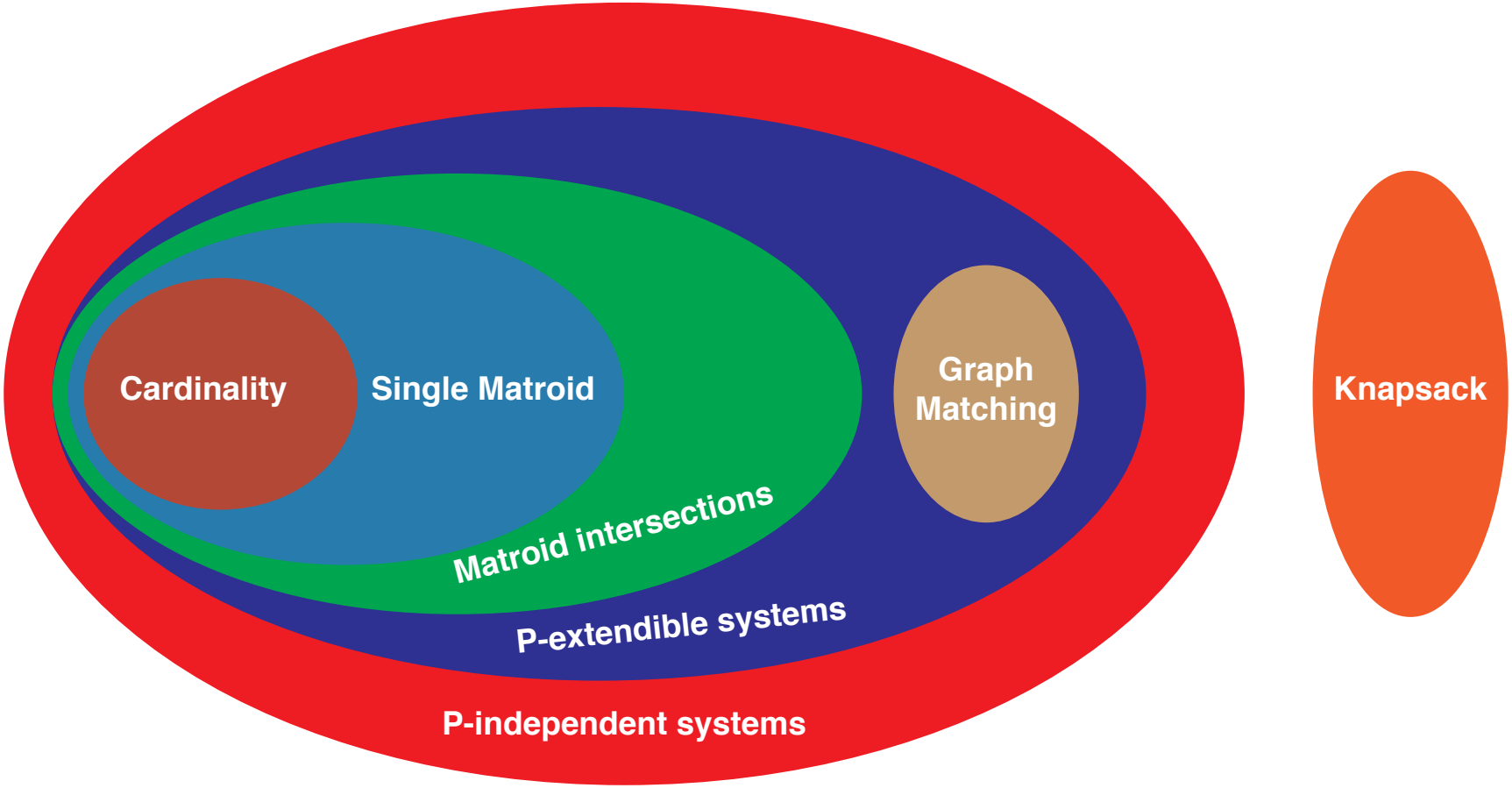


Figure idea from Amin Karbasi

Submodular Max Summary - From J. Vondrak

Monotone Maximization

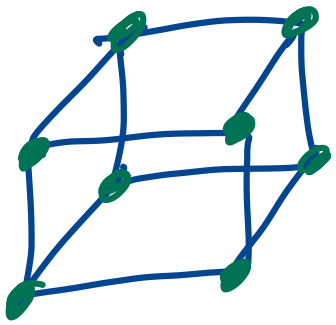
Constraint	Approximation	Hardness	Technique
$\lvert S\rvert \leq k$	$1-1/e$	$1-1/e$	greedy
matroid	$1-1/e$	$1-1/e$	multilinear ext.
$O(1)$ knapsacks	$1-1/e$	$1-1/e$	multilinear ext.
k matroids	$k+\epsilon$	$k/\log k$	local search
k matroids and $O(1)$ knapsacks	$O(k)$	$k/\log k$	multilinear ext.

Nonmonotone Maximization

Constraint	Approximation	Hardness	Technique
Unconstrained	$1/2$	$1/2$	combinatorial
matroid	$1/e$	0.48	multilinear ext.
$O(1)$ knapsacks	$1/e$	0.49	multilinear ext.
k matroids	$k+O(1)$	$k/\log k$	local search
k matroids and $O(1)$ knapsacks	$O(k)$	$k/\log k$	multilinear ext.

Continuous Extensions of Discrete Set Functions

- Any function $f : 2^V \to \mathbb{R}$ (equivalently $f : \{0,1\}^V \to \mathbb{R}$) can be extended to a continuous function in the sense $\tilde{f} : [0,1]^V \to \mathbb{R}$.



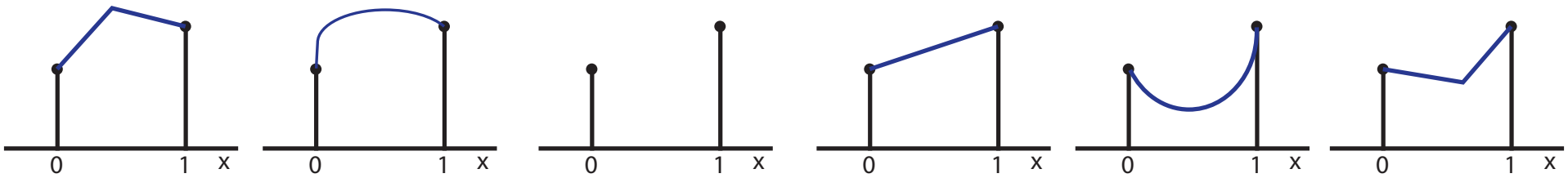
Continuous Extensions of Discrete Set Functions

- Any function $f : 2^V \to \mathbb{R}$ (equivalently $f : \{0,1\}^V \to \mathbb{R}$) can be extended to a continuous function in the sense $\tilde{f} : [0,1]^V \to \mathbb{R}$.
- This may be tight (i.e., $\tilde{f}(\mathbf{1}_A) = f(A)$ for all A). I.e., the extension $\tilde{f}$ coincides with f at the hypercube vertices.

Continuous Extensions of Discrete Set Functions

- Any function $f : 2^V \to \mathbb{R}$ (equivalently $f : \{0,1\}^V \to \mathbb{R}$) can be extended to a continuous function in the sense $\tilde{f} : [0,1]^V \to \mathbb{R}$.
- This may be tight (i.e., $\tilde{f}(\mathbf{1}_A) = f(A)$ for all A). I.e., the extension $\tilde{f}$ coincides with f at the hypercube vertices.
- In fact, any such discrete function defined on the vertices of the n-D hypercube $\{0,1\}^n$ has a variety of both convex and concave extensions tight at the vertices (Crama & Hammer'11). Example $n = 1$,

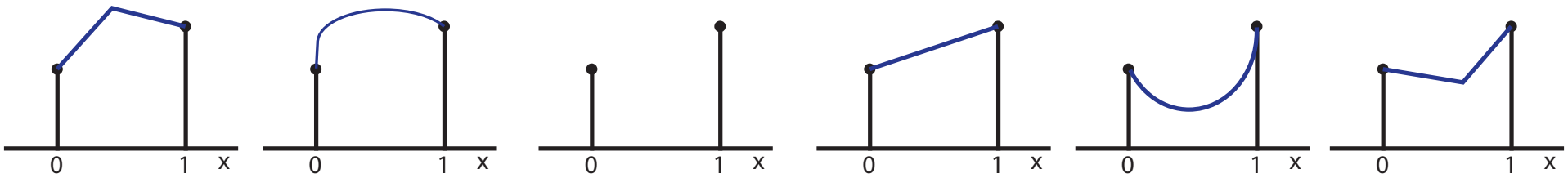
Concave Extensions $\tilde{f} : [0,1] \to \mathbb{R}$ | Discrete Function $f : \{0,1\}^V \to \mathbb{R}$ | Convex Extensions $\tilde{f} : [0,1] \to \mathbb{R}$



Continuous Extensions of Discrete Set Functions

- Any function $f : 2^V \to \mathbb{R}$ (equivalently $f : \{0,1\}^V \to \mathbb{R}$) can be extended to a continuous function in the sense $\tilde{f} : [0,1]^V \to \mathbb{R}$.
- This may be tight (i.e., $\tilde{f}(\mathbf{1}_A) = f(A)$ for all A). I.e., the extension $\tilde{f}$ coincides with f at the hypercube vertices.
- In fact, any such discrete function defined on the vertices of the n-D hypercube $\{0,1\}^n$ has a variety of both convex and concave extensions tight at the vertices (Crama & Hammer'11). Example $n = 1$,

Concave Extensions $\tilde{f} : [0,1] \to \mathbb{R}$ | Discrete Function $f : \{0,1\}^V \to \mathbb{R}$ | Convex Extensions $\tilde{f} : [0,1] \to \mathbb{R}$

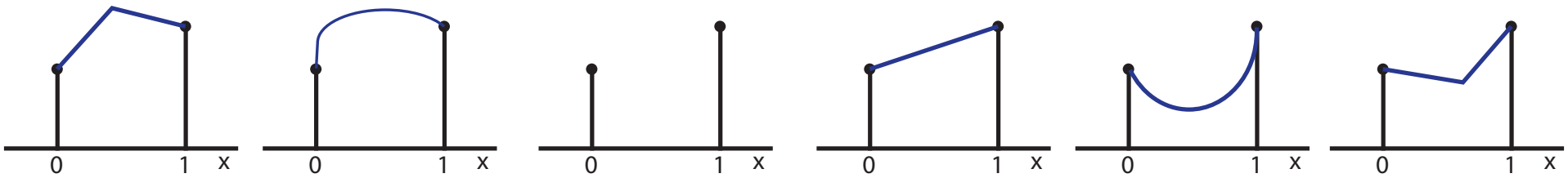


- Since there are an exponential number of vertices $\{0,1\}^n$, important questions regarding such extensions is:

Continuous Extensions of Discrete Set Functions

- Any function $f : 2^V \to \mathbb{R}$ (equivalently $f : \{0,1\}^V \to \mathbb{R}$) can be extended to a continuous function in the sense $\tilde{f} : [0,1]^V \to \mathbb{R}$.
- This may be tight (i.e., $\tilde{f}(\mathbf{1}_A) = f(A)$ for all A). I.e., the extension $\tilde{f}$ coincides with f at the hypercube vertices.
- In fact, any such discrete function defined on the vertices of the n-D hypercube $\{0,1\}^n$ has a variety of both convex and concave extensions tight at the vertices (Crama & Hammer'11). Example $n = 1$,

Concave Extensions $\tilde{f} : [0,1] \to \mathbb{R}$ | Discrete Function $f : \{0,1\}^V \to \mathbb{R}$ | Convex Extensions $\tilde{f} : [0,1] \to \mathbb{R}$

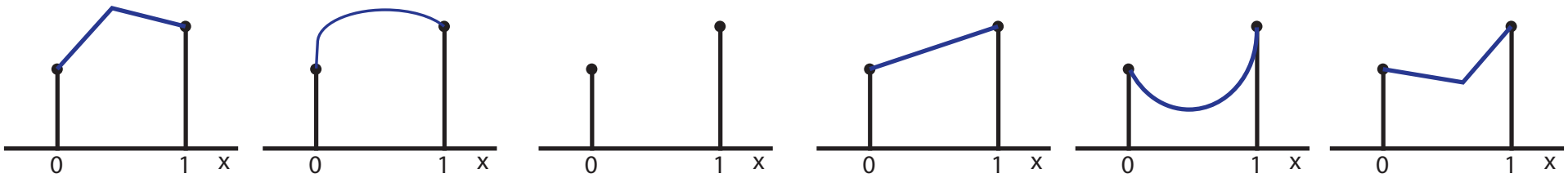


- Since there are an exponential number of vertices $\{0,1\}^n$, important questions regarding such extensions is:
  1. When are they computationally feasible to obtain or estimate?

Continuous Extensions of Discrete Set Functions

- Any function $f : 2^V \to \mathbb{R}$ (equivalently $f : \{0,1\}^V \to \mathbb{R}$) can be extended to a continuous function in the sense $\tilde{f} : [0,1]^V \to \mathbb{R}$.
- This may be tight (i.e., $\tilde{f}(\mathbf{1}_A) = f(A)$ for all A). I.e., the extension $\tilde{f}$ coincides with f at the hypercube vertices.
- In fact, any such discrete function defined on the vertices of the n-D hypercube $\{0,1\}^n$ has a variety of both convex and concave extensions tight at the vertices (Crama & Hammer'11). Example $n = 1$,

Concave Extensions $\tilde{f} : [0,1] \to \mathbb{R}$ Discrete Function $f : \{0,1\}^V \to \mathbb{R}$ Convex Extensions $\tilde{f} : [0,1] \to \mathbb{R}$

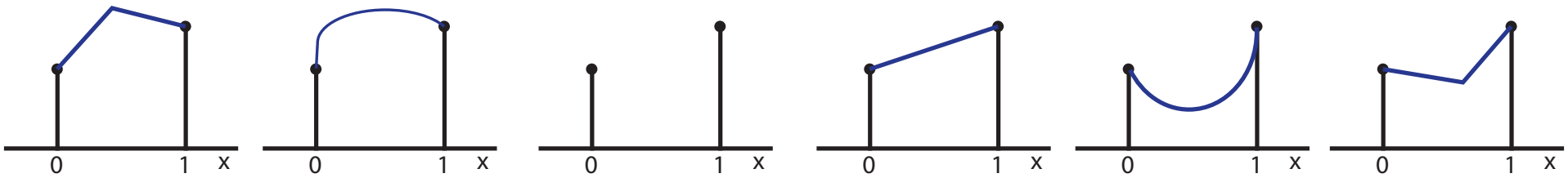


- Since there are an exponential number of vertices $\{0,1\}^n$, important questions regarding such extensions is:
  1. When are they computationally feasible to obtain or estimate?
  2. When do they have nice mathematical properties?

Continuous Extensions of Discrete Set Functions

- Any function $f : 2^V \to \mathbb{R}$ (equivalently $f : \{0,1\}^V \to \mathbb{R}$) can be extended to a continuous function in the sense $\tilde{f} : [0,1]^V \to \mathbb{R}$.
- This may be tight (i.e., $\tilde{f}(\mathbf{1}_A) = f(A)$ for all A). I.e., the extension $\tilde{f}$ coincides with f at the hypercube vertices.
- In fact, any such discrete function defined on the vertices of the n-D hypercube $\{0,1\}^n$ has a variety of both convex and concave extensions tight at the vertices (Crama & Hammer'11). Example $n = 1$,

Concave Extensions $\tilde{f} : [0,1] \to \mathbb{R}$ | Discrete Function $f : \{0,1\}^V \to \mathbb{R}$ | Convex Extensions $\tilde{f} : [0,1] \to \mathbb{R}$



- Since there are an exponential number of vertices $\{0,1\}^n$, important questions regarding such extensions is:
  1. When are they computationally feasible to obtain or estimate?
  2. When do they have nice mathematical properties?
  3. When are they useful for something practical? (4) which ~~one~~ extension are the useful ones?

Def: Convex Envelope of a function

- Given any function $h : \mathcal{D}_h \to \mathbb{R}$, where $\mathcal{D}_h \subseteq \mathbb{R}^n$, define the new function $\check{h} : \mathbf{R}^n \to \mathbb{R}$ via:

$$\check{h}(x) = \sup \{g(x) : g \text{ is convex } \& \ g(y) \leq h(y), \forall y \in \mathcal{D}_h\} \tag{16.1}$$

Def: Convex Envelope of a function

- Given any function $h : \mathcal{D}_h \to \mathbb{R}$, where $\mathcal{D}_h \subseteq \mathbb{R}^n$, define the new function $\check{h} : \mathbf{R}^n \to \mathbb{R}$ via:

$$\check{h}(x) = \sup \{g(x) : g \text{ is convex } \& \ g(y) \leq h(y), \forall y \in \mathcal{D}_h\} \tag{16.1}$$

- I.e., (1) $\check{h}(x)$ is convex, (2) $\check{h}(x) \leq h(x), \forall x$, and (3) if $g(x)$ is any convex function having the property that $g(x) \leq h(x), \forall x$, then $g(x) \leq \check{h}(x)$.

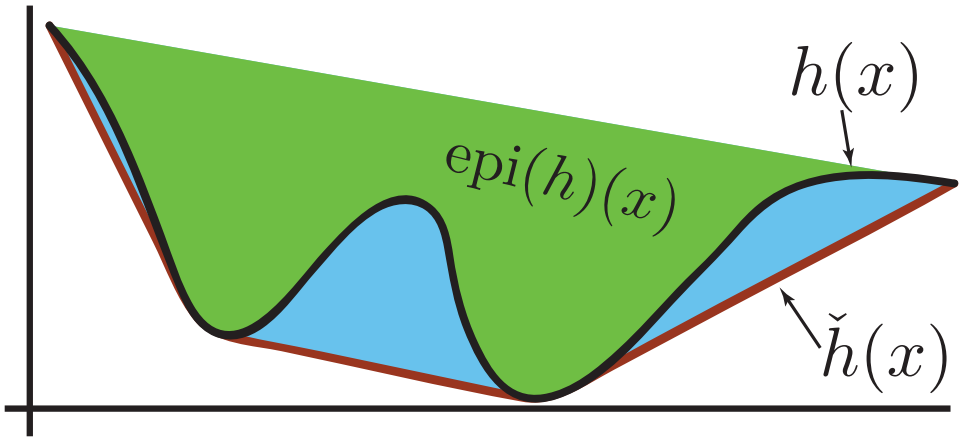
Def: Convex Envelope of a function

- Given any function $h : \mathcal{D}_h \to \mathbb{R}$, where $\mathcal{D}_h \subseteq \mathbb{R}^n$, define the new function $\check{h} : \mathbf{R}^n \to \mathbb{R}$ via:

$$\check{h}(x) = \sup \{g(x) : g \text{ is convex } \& \ g(y) \leq h(y), \forall y \in \mathcal{D}_h\} \tag{16.1}$$

- I.e., (1) $\check{h}(x)$ is convex, (2) $\check{h}(x) \leq h(x), \forall x$, and (3) if $g(x)$ is any convex function having the property that $g(x) \leq h(x), \forall x$, then $g(x) \leq \check{h}(x)$.
- Alternatively,

$$\check{h}(x) = \inf \{t : (x, t) \in \text{convexhull}(\text{epigraph}(h))\} \tag{16.2}$$

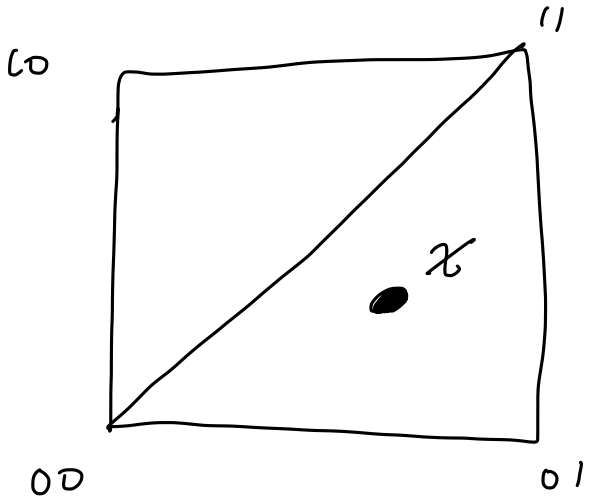


Convex Closure of Discrete Set Functions

- Given set function $f : 2^V \to \mathbb{R}$, an arbitrary (i.e., not necessarily submodular nor supermodular) set function, define a function $\check{f} : [0,1]^V \to \mathbb{R}$, as

$$\check{f}(x) = \min_{p \in \triangle^n(x)} \sum_{S \subseteq V} p_S f(S) \tag{16.3}$$

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



$x = p_{00} 1_{\phi} + p_{01} 1_{\{v_2\}}$
$+ p_{10} 1_{\{v_1\}}$

Convex Closure of Discrete Set Functions

- Given set function $f : 2^V \to \mathbb{R}$, an arbitrary (i.e., not necessarily submodular nor supermodular) set function, define a function $\check{f} : [0,1]^V \to \mathbb{R}$, as

$$\check{f}(x) = \min_{p \in \triangle^n(x)} \overbrace{\sum_{S \subseteq V} p_S f(S)}^{2^{|V|} \text{ terms in the sum.}} \tag{16.3}$$

$x \in [0,1]^V$

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

- Hence, $\triangle^n(x)$ is the set of all probability distributions over the 2^n vertices of the hypercube, and where the expected value of the characteristic vectors of those points is equal to x, i.e., for any $p \in \triangle^n(x)$, $E_{S \sim p}(\mathbf{1}_S) = \sum_{S \subseteq V} p_S \mathbf{1}_S = x$.

Convex Closure of Discrete Set Functions

- Given set function $f : 2^V \to \mathbb{R}$, an arbitrary (i.e., not necessarily submodular nor supermodular) set function, define a function $\check{f} : [0,1]^V \to \mathbb{R}$, as

$$\check{f}(x) = \min_{p \in \triangle^n(x)} \sum_{S \subseteq V} p_S f(S) \tag{16.3}$$

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

- Hence, $\triangle^n(x)$ is the set of all probability distributions over the 2^n vertices of the hypercube, and where the expected value of the characteristic vectors of those points is equal to x, i.e., for any $p \in \triangle^n(x)$, $E_{S \sim p}(\mathbf{1}_S) = \sum_{S \subseteq V} p_S \mathbf{1}_S = x$.
- Hence, $\check{f}(x) = \min_{p \in \triangle^n(x)} E_{S \sim p}[f(S)]$

Convex Closure of Discrete Set Functions

- Given set function $f : 2^V \to \mathbb{R}$, an arbitrary (i.e., not necessarily submodular nor supermodular) set function, define a function $\check{f} : [0,1]^V \to \mathbb{R}$, as

$$\check{f}(x) = \min_{p \in \triangle^n(x)} \sum_{S \subseteq V} p_S f(S) \tag{16.3}$$

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

- Hence, $\triangle^n(x)$ is the set of all probability distributions over the 2^n vertices of the hypercube, and where the expected value of the characteristic vectors of those points is equal to x, i.e., for any $p \in \triangle^n(x)$, $E_{S \sim p}(\mathbf{1}_S) = \sum_{S \subseteq V} p_S \mathbf{1}_S = x$.
- Hence, $\check{f}(x) = \min_{p \in \triangle^n(x)} E_{S \sim p}[f(S)]$
- ~~So far, as far as we know~~, this is ~~just~~ **a** convex extension. Does it have any special properties? we will show that

Convex Closure of Discrete Set Functions

- Given, $\check{f}(x) = \min_{p \in \triangle^n(x)} E_{S \sim p}[f(S)]$, we can show:

Convex Closure of Discrete Set Functions

- Given, $\check{f}(x) = \min_{p \in \triangle^n(x)} E_{S \sim p}[f(S)]$, we can show:
  1. that $\check{f}$ is tight (i.e., $\forall S \subseteq V$, we have $\check{f}(\mathbf{1}_S) = f(S)$).

Convex Closure of Discrete Set Functions

- Given, $\check{f}(x) = \min_{p \in \triangle^n(x)} E_{S \sim p}[f(S)]$, we can show:
  1. that $\check{f}$ is tight (i.e., $\forall S \subseteq V$, we have $\check{f}(\mathbf{1}_S) = f(S)$).
  2. that $\check{f}$ is convex (and consequently, that any arbitrary set function has a tight convex extension).

tight concave extension

Convex Closure of Discrete Set Functions

- Given, $\check{f}(x) = \min_{p \in \triangle^n(x)} E_{S \sim p}[f(S)]$, we can show:
  1. that $\check{f}$ is tight (i.e., $\forall S \subseteq V$, we have $\check{f}(\mathbf{1}_S) = f(S)$).
  2. that $\check{f}$ is convex (and consequently, that any arbitrary set function has a tight convex extension).
  3. that the convex closure $\check{f}$ is the convex envelope of the function defined only on the hypercube vertices, and that takes value $f(S)$ at $\mathbf{1}_S$.

$\mathcal{D}_h = \{0,1\}^V$

Convex Closure of Discrete Set Functions

- Given, $\check{f}(x) = \min_{p \in \triangle^n(x)} E_{S \sim p}[f(S)]$, we can show:
  1. that $\check{f}$ is tight (i.e., $\forall S \subseteq V$, we have $\check{f}(\mathbf{1}_S) = f(S)$).
  2. that $\check{f}$ is convex (and consequently, that any arbitrary set function has a tight convex extension).
  3. that the convex closure $\check{f}$ is the convex envelope of the function defined only on the hypercube vertices, and that takes value $f(S)$ at $\mathbf{1}_S$.
  4. the definition of the Lovász extension of a set function, and that $\check{f}$ is the Lovász extension iff f is submodular.

Tightness of Convex Closure

Lemma 16.4.1

$\forall A \subseteq V$, *we have* $\check{f}(\mathbf{1}_A) = f(A)$.

Proof.

- Define p^x to be an achieving argmin in $\check{f}(x) = \min_{p \in \triangle^n(x)} E_{S \sim p}[f(S)]$.

Tightness of Convex Closure

Lemma 16.4.1

$\forall A \subseteq V$, *we have* $\check{f}(\mathbf{1}_A) = f(A)$.

Proof.

- Define p^x to be an achieving argmin in $\check{f}(x) = \min_{p \in \triangle^n(x)} E_{S \sim p}[f(S)]$.
- Take an arbitrary A, so that $\mathbf{1}_A = \sum_{S \subseteq V} p_S^{\mathbf{1}_A} \mathbf{1}_S$.

what are properties of $\{p_S\}_{S \subseteq V}$?

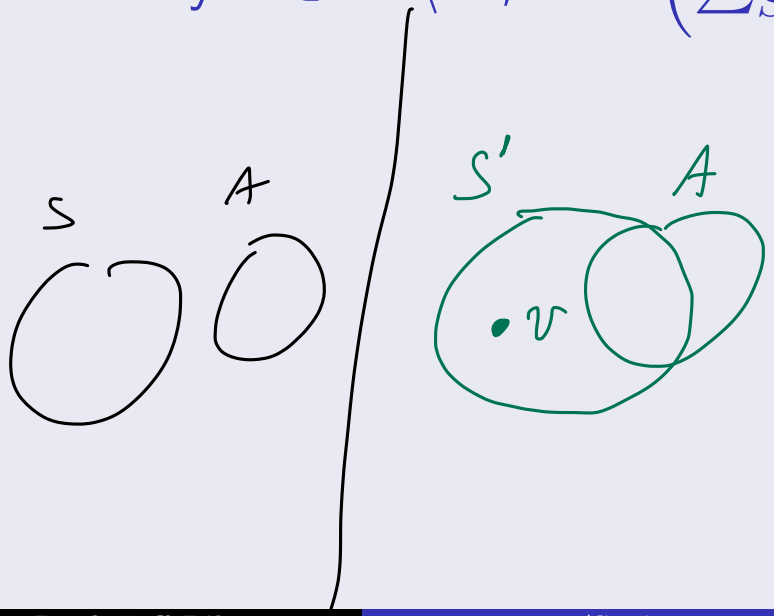
Tightness of Convex Closure

Lemma 16.4.1

$\forall A \subseteq V$, *we have* $\check{f}(\mathbf{1}_A) = f(A)$.

Proof.

- Define p^x to be an achieving argmin in $\check{f}(x) = \min_{p \in \triangle^n(x)} E_{S \sim p}[f(S)]$.
- Take an arbitrary A, so that $\mathbf{1}_A = \sum_{S \subseteq V} p_S^{\mathbf{1}_A} \mathbf{1}_S$.
- Suppose $\exists S'$ with $S' \setminus A \neq 0$ having $p_{S'}^{\mathbf{1}_A} > 0$. This would mean, for any $v \in S' \setminus A$, that $\left(\sum_S p_S^{\mathbf{1}_A} \mathbf{1}_S\right)(v) > 0$, a contradiction.



$\therefore$ All such S' have $p_{S'}^{\mathbf{1}_A} = 0$.

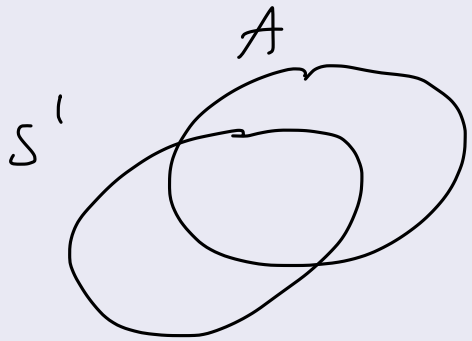
Tightness of Convex Closure

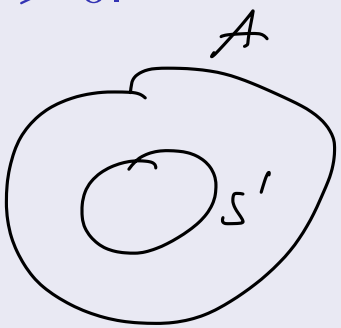
Lemma 16.4.1

$\forall A \subseteq V$, *we have* $\check{f}(\mathbf{1}_A) = f(A)$.

Proof.

- Define p^x to be an achieving argmin in $\check{f}(x) = \min_{p \in \triangle^n(x)} E_{S \sim p}[f(S)]$.
- Take an arbitrary A, so that $\mathbf{1}_A = \sum_{S \subseteq V} p_S^{\mathbf{1}_A} \mathbf{1}_S$.
- Suppose $\exists S'$ with $S' \setminus A \neq 0$ having $p_{S'}^{\mathbf{1}_A} > 0$. This would mean, for any $v \in S' \setminus A$, that $\left(\sum_S p_S^{\mathbf{1}_A} \mathbf{1}_S\right)(v) > 0$, a contradiction.
- Suppose $\exists S'$ s.t. $A \setminus S' \neq \emptyset$ with $p_{S'}^{\mathbf{1}_A} > 0$.





□

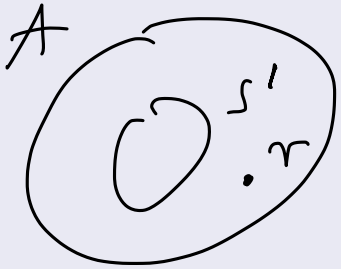
Tightness of Convex Closure

Lemma 16.4.1

$\forall A \subseteq V$, *we have* $\check{f}(\mathbf{1}_A) = f(A)$.

Proof.

- Define p^x to be an achieving argmin in $\check{f}(x) = \min_{p \in \triangle^n(x)} E_{S \sim p}[f(S)]$.
- Take an arbitrary A, so that $\mathbf{1}_A = \sum_{S \subseteq V} p_S^{\mathbf{1}_A} \mathbf{1}_S$.
- Suppose $\exists S'$ with $S' \setminus A \neq 0$ having $p_{S'}^{\mathbf{1}_A} > 0$. This would mean, for any $v \in S' \setminus A$, that $\left(\sum_S p_S^{\mathbf{1}_A} \mathbf{1}_S\right)(v) > 0$, a contradiction.
- Suppose $\exists S'$ s.t. $A \setminus S' \neq \emptyset$ with $p_{S'}^{\mathbf{1}_A} > 0$.
- Then, for any $v \in A \setminus S'$, consider below leading to a contradiction



$$\underbrace{p_{S'} \mathbf{1}_{S'}}_{>0} + \underbrace{\sum_{\substack{S \subseteq A \\ S \neq S'}} p_S \mathbf{1}_S}_{\text{can't sum to 1}} \Rightarrow \left(\sum_{\substack{S \subseteq A \\ S \neq S'}} p_s \mathbf{1}_S\right)(v) < 1 \tag{16.4}$$

$\therefore p_{S'}^{1_A} = 0$

$\therefore p_A^{1_A} = 1$

I.e., $v \in A$ so it must get value 1, but since $v \notin S'$, v is deficient. □

Convexity of the Convex Closure

Lemma 16.4.2

$\check{f}(x) = \min_{p \in \triangle^n(x)} E_{S \sim p}[f(S)]$ *is convex in* $[0,1]^V$.

Proof.

- Let $x, y \in [0,1]^V$, $0 \leq \lambda \leq 1$, and $z = \lambda x + (1-\lambda)y$, then

$$\lambda \check{f}(x) + (1-\lambda)\check{f}(y) = \lambda \sum_S p_S^x f(S) + (1-\lambda) \sum_S p_S^y f(S) \quad (16.5)$$

$$= \sum_S (\lambda p_S^x + (1-\lambda) p_S^y) f(S) \quad (16.6)$$

$$= \sum_S p_S^{z'} f(S) \geq \min_{p \in \triangle^n(z)} E_{S \sim p}[f(S)] \quad (16.7)$$

$$= \check{f}(z) = \check{f}(\lambda x + (1-\lambda)y) \quad (16.8)$$

Convexity of the Convex Closure

Lemma 16.4.2

$\check{f}(x) = \min_{p \in \triangle^n(x)} E_{S \sim p}[f(S)]$ *is convex in* $[0,1]^V$.

Proof.

- Let $x, y \in [0,1]^V$, $0 \leq \lambda \leq 1$, and $z = \lambda x + (1-\lambda)y$, then

$$\lambda \check{f}(x) + (1-\lambda)\check{f}(y) = \lambda \sum_S p_S^x f(S) + (1-\lambda) \sum_S p_S^y f(S) \quad (16.5)$$

$$= \sum_S (\lambda p_S^x + (1-\lambda) p_S^y) f(S) \quad (16.6)$$

$$= \sum_S p_S^{z'} f(S) \geq \min_{p \in \triangle^n(z)} E_{S \sim p}[f(S)] \quad (16.7)$$

$$= \check{f}(z) = \check{f}(\lambda x + (1-\lambda)y) \quad (16.8)$$

- Note that $p_S^{z'} = \lambda p_S^x + (1-\lambda) p_S^y$ and is feasible in the min since $\sum_S p_S^{z'} = 1$, $p_S^{z'} \geq 0$ and $\sum_S p_S^{z'} \mathbf{1}_S = z$. $p_S^{z'} \in \triangle^n(z)$

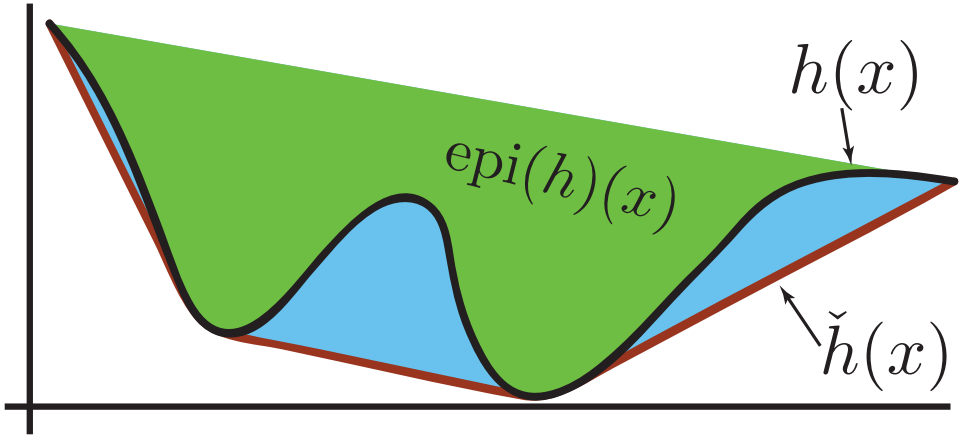
Def: Convex Envelope of a function

- Given any function $h : \mathcal{D}_h \to \mathbb{R}$, where $\mathcal{D}_h \subseteq \mathbb{R}^n$, define the new function $\check{h} : \mathbf{R}^n \to \mathbb{R}$ via:

$$\check{h}(x) = \sup \{g(x) : g \text{ is convex } \& \ g(y) \leq h(y), \forall y \in \mathcal{D}_h\} \tag{16.1}$$

- I.e., (1) $\check{h}(x)$ is convex, (2) $\check{h}(x) \leq h(x), \forall x$, and (3) if $g(x)$ is any convex function having the property that $g(x) \leq h(x), \forall x$, then $g(x) \leq \check{h}(x)$.
- Alternatively,

$$\check{h}(x) = \inf \{t : (x, t) \in \text{convexhull}(\text{epigraph}(h))\} \tag{16.2}$$



Convex Closure is the Convex Envelope

Lemma 16.4.3

$\check{f}(x) = \min_{p \in \triangle^n(x)} E_{S \sim p}[f(S)]$ *is the convex envelope.*

Proof.

- Suppose $\exists$ a convex $\bar{f}$ with $\bar{f}(\mathbf{1}_A) = f(A) = \check{f}(\mathbf{1}_A), \forall A \subseteq V$ and $\exists x \in [0,1]^V$ s.t. $\bar{f}(x) > \check{f}(x)$.
- Define p^x to be an achieving argmin in $\check{f}(x) = \min_{p \in \triangle^n(x)} E_{S \sim p}[f(S)]$. Hence, we have $x = \sum_S p_S^x \mathbf{1}_S$. Thus

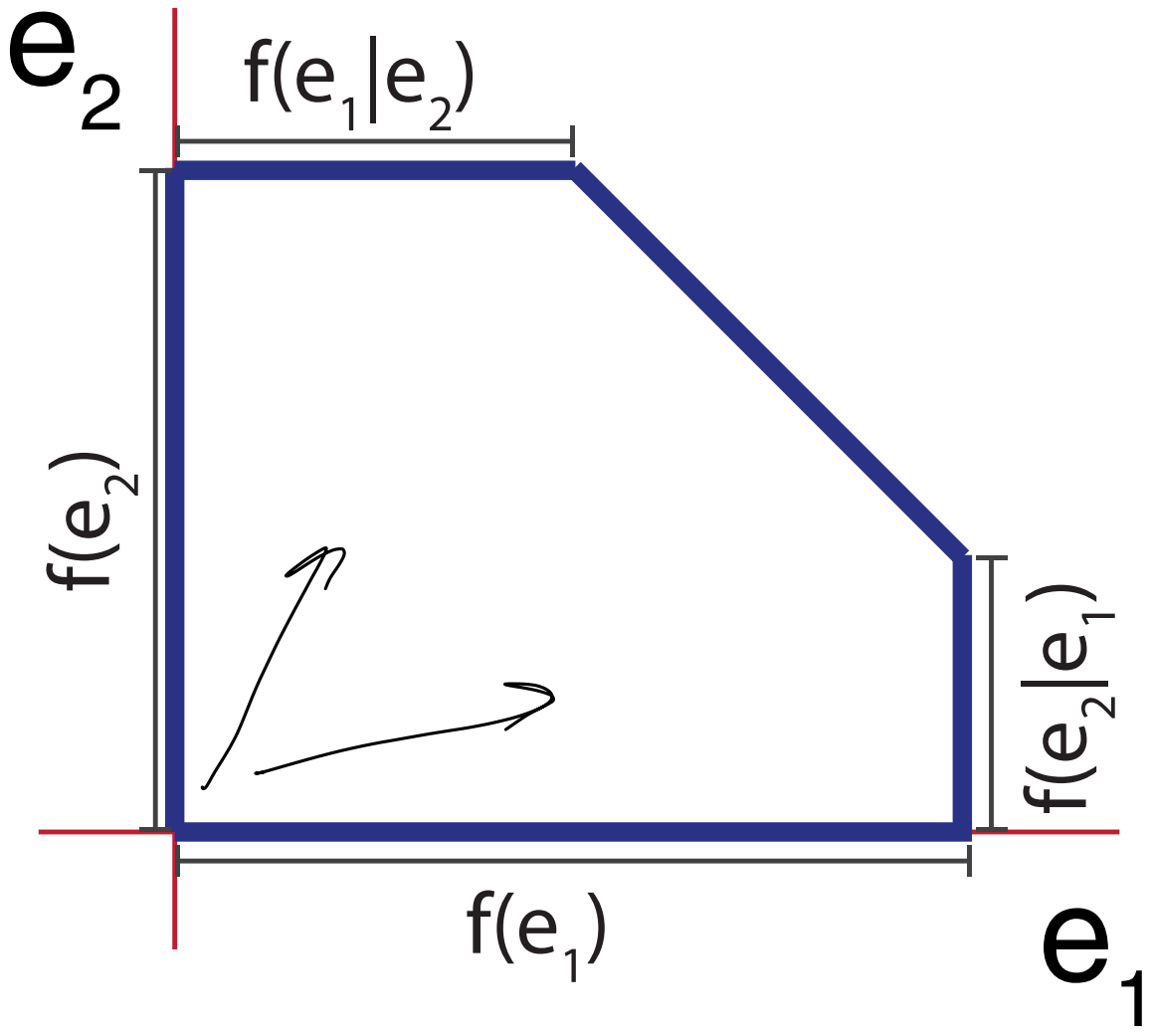
$$\check{f}(x) = \sum_S p_S^x f(S) = \sum_S p_S^x \bar{f}(\mathbf{1}_S) \tag{16.9}$$

$$< \bar{f}(x) = \bar{f}(\sum_S p_S^x \mathbf{1}_S) \tag{16.10}$$

but the inequality contradicts the convexity of $\bar{f}$.

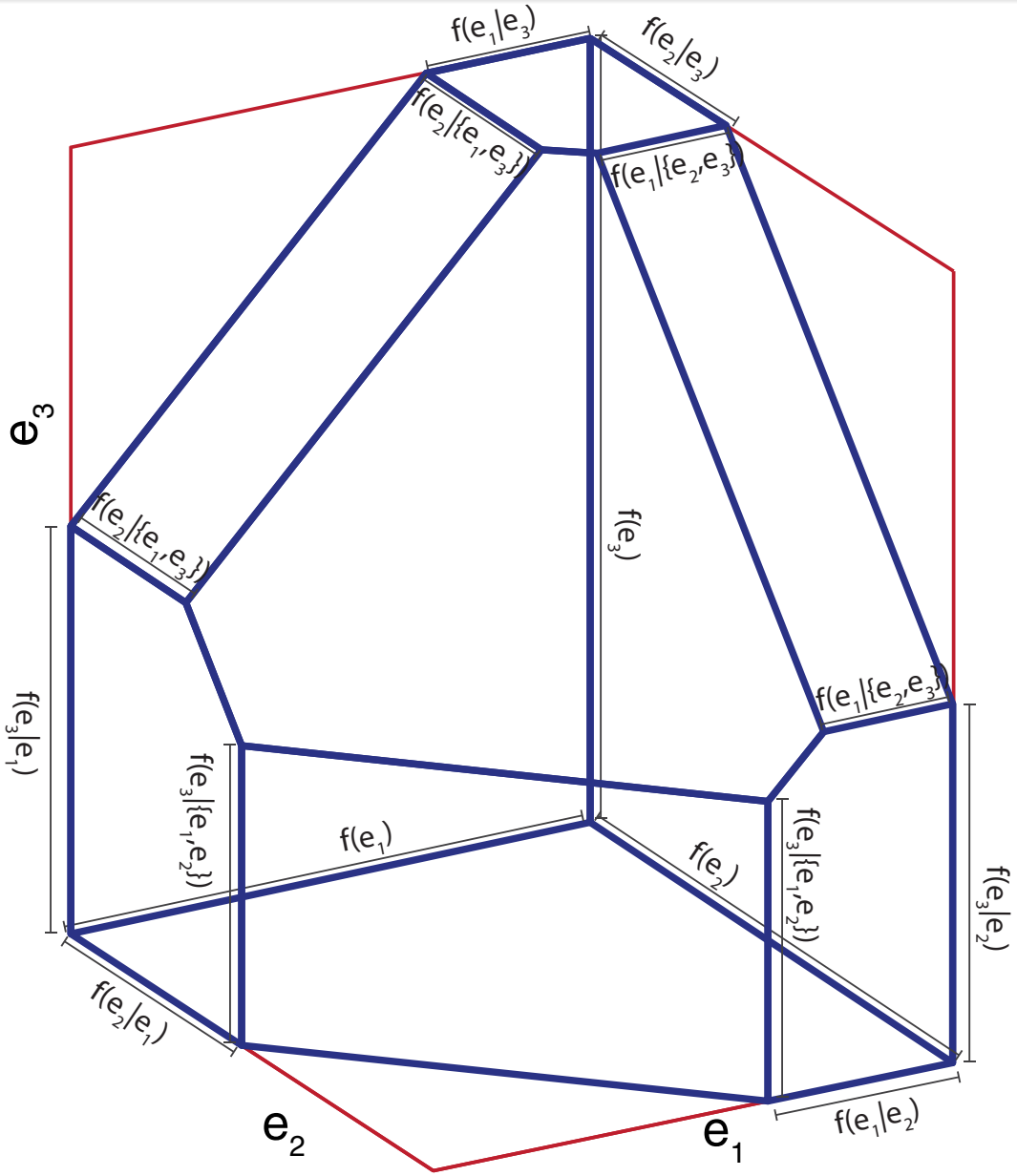
Polymatroid with labeled edge lengths

- Recall $f(e|A) = f(A+e) - f(A)$
- Notice how submodularity, $f(e|B) \leq f(e|A)$ for $A \subseteq B$, defines the shape of the polytope.
- In fact, we have strictness here $f(e|B) < f(e|A)$ for $A \subset B$.
- Also, consider how the greedy algorithm proceeds along the edges of the polytope.



Polymatroid with labeled edge lengths

- Recall $f(e|A) = f(A+e) - f(A)$
- Notice how submodularity, $f(e|B) \leq f(e|A)$ for $A \subseteq B$, defines the shape of the polytope.
- In fact, we have strictness here $f(e|B) < f(e|A)$ for $A \subset B$.
- Also, consider how the greedy algorithm proceeds along the edges of the polytope.



Optimization over P_f

- Consider the following optimization. Given $w \in \mathbb{R}^E$,

$$\begin{aligned} \text{maximize} \quad & w^\intercal x && (16.11a) \\ \text{subject to} \quad & x \in P_f && (16.11b) \end{aligned}$$

- Since P_f is down closed, if $\exists e \in E$ with $w(e) < 0$ then the solution above is unboundedly large. Hence, assume $w \in \mathbb{R}_+^E$.
- In a future lecture, we will see that any $x \in P_f$ with $x \notin B_f$ is dominated by $x \leq y \in B_f$ which can only increase $w^\intercal x \leq w^\intercal y$ when $w \in \mathbb{R}_+^E$.
- Hence, the problem is equivalent to: given $w \in \mathbb{R}_+^E$,

$$\begin{aligned} \text{maximize} \quad & w^\intercal x && (16.12a) \\ \text{subject to} \quad & x \in B_f && (16.12b) \end{aligned}$$

- Moreover, we can have $w \in \mathbb{R}^E$ if we insist on $x \in B_f$.

A continuous extension of f

- Consider again optimization problem. Given $w \in \mathbb{R}^E$,

$$\begin{aligned} \text{maximize} \quad & w^\intercal x && (16.13a) \\ \text{subject to} \quad & x \in B_f && (16.13b) \end{aligned}$$

A continuous extension of f

- Consider again optimization problem. Given $w \in \mathbb{R}^E$,

$$\begin{aligned} \text{maximize} \quad & w^\intercal x && (16.13a) \\ \text{subject to} \quad & x \in B_f && (16.13b) \end{aligned}$$

- We may consider this optimization problem a function $\breve{f} : \mathbb{R}^E \to \mathbb{R}$ of $w \in \mathbb{R}^E$, defined as:

\breve f

$$\breve{f}(w) = \max(wx : x \in B_f) \tag{16.14}$$

A continuous extension of f

- Consider again optimization problem. Given $w \in \mathbb{R}^E$,

$$\begin{aligned} \text{maximize} \quad & w^\intercal x && (16.13a) \\ \text{subject to} \quad & x \in B_f && (16.13b) \end{aligned}$$

- We may consider this optimization problem a function $\breve{f} : \mathbb{R}^E \to \mathbb{R}$ of $w \in \mathbb{R}^E$, defined as:

$$\breve{f}(w) = \max(wx : x \in B_f) \tag{16.14}$$

- Note that this is $\breve{f}(w)$ which is a distinct notation from the convex closure $\breve{f}(x)$ we defined before.

A continuous extension of f

- Consider again optimization problem. Given $w \in \mathbb{R}^E$,

$$\begin{aligned} \text{maximize} \quad & w^\intercal x && (16.13a) \\ \text{subject to} \quad & x \in B_f && (16.13b) \end{aligned}$$

- We may consider this optimization problem a function $\breve{f} : \mathbb{R}^E \to \mathbb{R}$ of $w \in \mathbb{R}^E$, defined as:

$$\breve{f}(w) = \max(wx : x \in B_f) \tag{16.14}$$

- Note that this is $\breve{f}(w)$ which is a distinct notation from the convex closure $\check{f}(x)$ we defined before.
- Hence, for any w, from the solution to the above theorem (as we have seen), we can compute the value of this function using Edmond's greedy algorithm.

Recall Edmond's Theorem: The Greedy Algorithm

- Recall, Edmonds proved that the solution to $\breve{f}(w) = \max(wx : x \in B_f)$ is solved by the greedy algorithm iff f is submodular.
- In particular, sort choose element order $(e_1, e_2, \dots, e_m)$ based on decreasing w,so that $w(e_1) \geq w(e_2) \geq \cdots \geq w(e_m)$.
- Define the chain with i^{th} element $E_i = \{e_1, e_2, \dots, e_i\}$.
- Define a vector $x^* \in \mathbb{R}^V$ where element e_i has value $x(e_i) = f(e_i | E_{i-1})$ for all $i \in V$.
- Then $\langle w, x^* \rangle = \max(wx : x \in B_f)$

Theorem 16.5.1 (Edmonds)

If $f : 2^E \to \mathbb{R}_+$ is given, and B_f is a polytope in $\mathbb{R}^E$ of the form $B_f = \left\{x \in \mathbb{R}^E : x(A) \leq f(A), \forall A \subseteq E, x(E) = f(E)\right\}$, then the greedy solution to the problem $\max(w^\intercal x : x \in B_f)$ is $\forall w$ optimum iff f is monotone non-decreasing submodular (i.e., iff the corresponding P_f is a polymatroid).

Greedy-based continuous extension of submodular f

- That is, given a submodular function f, a $w \in \mathbb{R}^E$, choose element order $(e_1, e_2, \ldots, e_m)$ based on decreasing w,so that $w(e_1) \geq w(e_2) \geq \cdots \geq w(e_m)$.

Greedy-based continuous extension of submodular f

- That is, given a submodular function f, a $w \in \mathbb{R}^E$, choose element order $(e_1, e_2, \ldots, e_m)$ based on decreasing w,so that $w(e_1) \geq w(e_2) \geq \cdots \geq w(e_m)$.
- Define chain $\emptyset \triangleq E_0 \subset E_1 \subset E_2 \subset \cdots \subset E_m = E$ based on w, so the ith element of this change has $E_i = \{e_1, e_2, \ldots, e_i\}$.

Greedy-based continuous extension of submodular f

- That is, given a submodular function f, a $w \in \mathbb{R}^E$, choose element order $(e_1, e_2, \ldots, e_m)$ based on decreasing w,so that $w(e_1) \geq w(e_2) \geq \cdots \geq w(e_m)$.
- Define chain $\emptyset \triangleq E_0 \subset E_1 \subset E_2 \subset \cdots \subset E_m = E$ based on w, so the ith element of this change has $E_i = \{e_1, e_2, \ldots, e_i\}$.

We have

$\breve{f}(w)$

Greedy-based continuous extension of submodular f

- That is, given a submodular function f, a $w \in \mathbb{R}^E$, choose element order $(e_1, e_2, \ldots, e_m)$ based on decreasing w,so that $w(e_1) \geq w(e_2) \geq \cdots \geq w(e_m)$.
- Define chain $\emptyset \triangleq E_0 \subset E_1 \subset E_2 \subset \cdots \subset E_m = E$ based on w, so the ith element of this change has $E_i = \{e_1, e_2, \ldots, e_i\}$.

We have

$$\breve{f}(w) = \max(wx : x \in B_f) \tag{16.15}$$

Greedy-based continuous extension of submodular f

- That is, given a submodular function f, a $w \in \mathbb{R}^E$, choose element order $(e_1, e_2, \ldots, e_m)$ based on decreasing w,so that $w(e_1) \geq w(e_2) \geq \cdots \geq w(e_m)$.
- Define chain $\emptyset \triangleq E_0 \subset E_1 \subset E_2 \subset \cdots \subset E_m = E$ based on w, so the ith element of this change has $E_i = \{e_1, e_2, \ldots, e_i\}$.

We have

$$\breve{f}(w) = \max(wx : x \in B_f) \tag{16.15}$$

$$= \sum_{i=1}^{m} w(e_i) f(e_i | E_{i-1}) = \sum_{i=1}^{m} w(e_i) x(e_i) \tag{16.16}$$

Greedy-based continuous extension of submodular f

- That is, given a submodular function f, a $w \in \mathbb{R}^E$, choose element order $(e_1, e_2, \ldots, e_m)$ based on decreasing w,so that $w(e_1) \geq w(e_2) \geq \cdots \geq w(e_m)$.
- Define chain $\emptyset \triangleq E_0 \subset E_1 \subset E_2 \subset \cdots \subset E_m = E$ based on w, so the ith element of this change has $E_i = \{e_1, e_2, \ldots, e_i\}$.

We have

$$\breve{f}(w) = \max(wx : x \in B_f) \tag{16.15}$$

$$= \sum_{i=1}^{m} w(e_i) f(e_i | E_{i-1}) = \sum_{i=1}^{m} w(e_i) x(e_i) \tag{16.16}$$

$$= \sum_{i=1}^{m} w(e_i)(f(E_i) - f(E_{i-1})) \tag{16.17}$$

Greedy-based continuous extension of submodular f

- That is, given a submodular function f, a $w \in \mathbb{R}^E$, choose element order $(e_1, e_2, \ldots, e_m)$ based on decreasing w,so that $w(e_1) \geq w(e_2) \geq \cdots \geq w(e_m)$.
- Define chain $\emptyset \triangleq E_0 \subset E_1 \subset E_2 \subset \cdots \subset E_m = E$ based on w, so the ith element of this change has $E_i = \{e_1, e_2, \ldots, e_i\}$.

We have

$$\breve{f}(w) = \max(wx : x \in B_f) \tag{16.15}$$

$$= \sum_{i=1}^{m} w(e_i) f(e_i | E_{i-1}) = \sum_{i=1}^{m} w(e_i) x(e_i) \tag{16.16}$$

$$= \sum_{i=1}^{m} w(e_i)(f(E_i) - f(E_{i-1})) \tag{16.17}$$

typically $=0$

$$= -w(e_1) f(E_0) + w(e_m) f(E_m) + \sum_{i=1}^{m-1} (w(e_i) - w(e_{i+1})) f(E_i) \tag{16.18}$$

Greedy-based continuous extension of submodular f

- Definition of the continuous extension, once again, for reference:

$$\breve{f}(w) = \max(wx : x \in B_f) \tag{16.19}$$

Greedy-based continuous extension of submodular f

- Definition of the continuous extension, once again, for reference:

$$\breve{f}(w) = \max(wx : x \in B_f) \tag{16.19}$$

- Therefore, if f is a submodular function, we can write

$$\breve{f}(w)$$

Greedy-based continuous extension of submodular f

- Definition of the continuous extension, once again, for reference:

$$\breve{f}(w) = \max(wx : x \in B_f) \tag{16.19}$$

- Therefore, if f is a submodular function, we can write

$$\breve{f}(w) = w(e_m)f(E_m) + \sum_{i=1}^{m-1}(w(e_i) - w(e_{i+1}))f(E_i) \tag{16.20}$$

Greedy-based continuous extension of submodular f

- Definition of the continuous extension, once again, for reference:

$$\breve{f}(w) = \max(wx : x \in B_f) \tag{16.19}$$

- Therefore, if f is a submodular function, we can write

$$\breve{f}(w) = w(e_m)f(E_m) + \sum_{i=1}^{m-1}(w(e_i) - w(e_{i+1}))f(E_i) \tag{16.20}$$

$$= \sum_{i=1}^{m} \lambda_i f(E_i) \tag{16.21}$$

Greedy-based continuous extension of submodular f

- Definition of the continuous extension, once again, for reference:

$$\breve{f}(w) = \max(wx : x \in B_f) \tag{16.19}$$

- Therefore, if f is a submodular function, we can write

$$\breve{f}(w) = w(e_m)f(E_m) + \sum_{i=1}^{m-1}(w(e_i) - w(e_{i+1}))f(E_i) \tag{16.20}$$

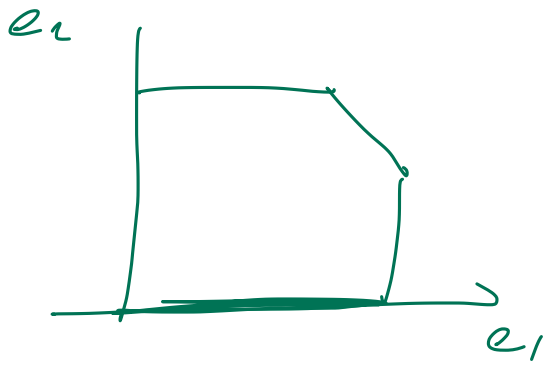
$$= \sum_{i=1}^{m}\lambda_i f(E_i) \tag{16.21}$$

where $\lambda_m = w(e_m)$ and otherwise $\lambda_i = w(e_i) - w(e_{i+1})$, where the elements are sorted descending according to w as before.

Greedy-based continuous extension: properties?

- So we go from f to B_f to $\breve{f}(w) = \max(wx : x \in B_f) = \sum_{i=1}^{m} \lambda_i f(E_i)$ when f is submodular.

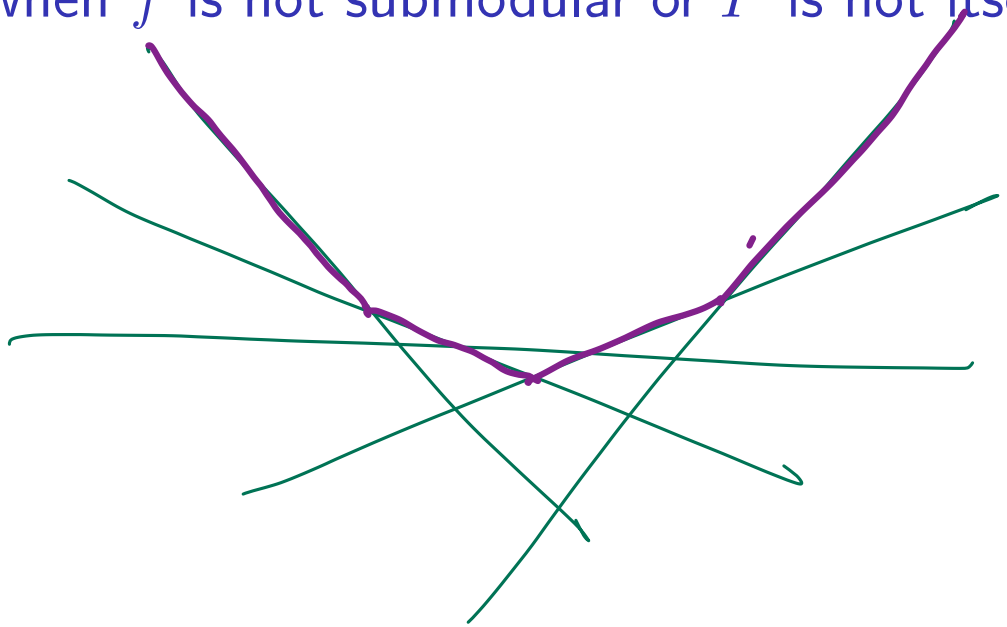
$\breve{f}(1_A) = f(A)$



$A = \{e_1\}$

Greedy-based continuous extension: properties?

- So we go from f to B_f to $\breve{f}(w) = \max(wx : x \in B_f) = \sum_{i=1}^{m} \lambda_i f(E_i)$ when f is submodular.
- Convex analysis $\Rightarrow$ $\breve{f}(w) = \max(wx : x \in P)$ is always convex in w for any set $P \subseteq R^E$, since a maximum of a set of linear functions (true even when f is not submodular or P is not itself a convex set).



Greedy-based continuous extension: properties?

- So we go from f to B_f to $\breve{f}(w) = \max(wx : x \in B_f) = \sum_{i=1}^{m} \lambda_i f(E_i)$ when f is submodular.
- Convex analysis $\Rightarrow \breve{f}(w) = \max(wx : x \in P)$ is always convex in w for any set $P \subseteq R^E$, since a maximum of a set of linear functions (true even when f is not submodular or P is not itself a convex set).
- What this says, though is that when f is submodular, then $\sum_{i=1}^{m} \lambda_i f(E_i)$ is convex.

Greedy-based continuous extension: properties?

- So we go from f to B_f to $\breve{f}(w) = \max(wx : x \in B_f) = \sum_{i=1}^{m} \lambda_i f(E_i)$ when f is submodular.
- Convex analysis $\Rightarrow \breve{f}(w) = \max(wx : x \in P)$ is always convex in w for any set $P \subseteq R^E$, since a maximum of a set of linear functions (true even when f is not submodular or P is not itself a convex set).
- What this says, though is that when f is submodular, then $\sum_{i=1}^{m} \lambda_i f(E_i)$ is convex.
- What can we say about the form $\sum_{i=1}^{m} \lambda_i f(E_i)$ in general? What is this particular form if f isn't even submodular?

Reminder on notation

- Recall, for any such $w \in \mathbb{R}^E$, we have

$$\begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} = \underbrace{(w_1 - w_2)}_{\lambda_1} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \underbrace{(w_2 - w_3)}_{\lambda_2} \begin{pmatrix} 1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} +$$

$\lambda_0 \geq 0$

$$\cdots + \underbrace{(w_{n-1} - w_n)}_{\lambda_{m-1}} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 0 \end{pmatrix} + \underbrace{(w_m)}_{\lambda_m} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{pmatrix} \tag{16.22}$$

- If we take w in decreasing order, then each coefficient of the vectors is non-negative (except possibly the last one, $\lambda_m = w_m$).
- We often take $w \in \mathbb{R}_+^V$ or even $w \in [0,1]^V$, making $\lambda_m \geq 0$.

More reminder on notation

Axis rename

- Again, order the elements based on w and define E_i based on decreasing order of w, giving $E_i = \{e_1, e_2, \ldots, e_i\}$ for $i = 0, \ldots, n$.
- Recall that

$$\mathbf{1}_{E_0} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \mathbf{1}_{E_1} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \mathbf{1}_{E_\ell} = \begin{pmatrix} \left.\begin{matrix} 1 \\ 1 \\ \vdots \\ 1 \end{matrix}\right\} \ell\times \\ \left.\begin{matrix} 0 \\ \\ 0 \\ \vdots \\ 0 \end{matrix}\right\} (n-\ell)\times \end{pmatrix}, \text{ etc.}$$

- Hence, from the previous and current slide, we have $w = \sum_{i=1}^{m} \lambda_i \mathbf{1}_{E_i}$

From $\breve{f}$ back to f, even when f is not submodular

- From the continuous $\breve{f}$, we can recover $f(A)$ for any $A \subseteq V$.

From $\breve{f}$ back to f, even when f is not submodular

- From the continuous $\breve{f}$, we can recover $f(A)$ for any $A \subseteq V$.
- Take $w = \mathbf{1}_A$ for some $A \subseteq E$, so w is vertex of the hypercube.

From $\breve{f}$ back to f, even when f is not submodular

- From the continuous $\breve{f}$, we can recover $f(A)$ for any $A \subseteq V$.
- Take $w = \mathbf{1}_A$ for some $A \subseteq E$, so w is vertex of the hypercube.
- Order and rename the elements of E in decreasing order of w so that $w(e_1) \geq w(e_2) \geq w(e_3) \geq \cdots \geq w(e_m)$.

$\sum_i \lambda_i f(E_i)$

From $\breve{f}$ back to f, even when f is not submodular

- From the continuous $\breve{f}$, we can recover $f(A)$ for any $A \subseteq V$.
- Take $w = \mathbf{1}_A$ for some $A \subseteq E$, so w is vertex of the hypercube.
- Order and rename the elements of E in decreasing order of w so that $w(e_1) \geq w(e_2) \geq w(e_3) \geq \cdots \geq w(e_m)$.
- This means

$$w = (w(e_1), w(e_2), \ldots, w(e_m)) = (\underbrace{1, 1, 1, \ldots, 1}_{|A| \text{ times}}, \underbrace{0, 0, \ldots, 0}_{m-|A| \text{ times}}) \quad (16.23)$$

so that $1_A(i) = 1$ if $i \leq |A|$, and $1_A(i) = 0$ otherwise.

$w(e_i) - w(e_{i+1})$

From $\breve{f}$ back to f, even when f is not submodular

- From the continuous $\breve{f}$, we can recover $f(A)$ for any $A \subseteq V$.
- Take $w = \mathbf{1}_A$ for some $A \subseteq E$, so w is vertex of the hypercube.
- Order and rename the elements of E in decreasing order of w so that $w(e_1) \geq w(e_2) \geq w(e_3) \geq \cdots \geq w(e_m)$.
- This means

$$w = (w(e_1), w(e_2), \ldots, w(e_m)) = (\underbrace{1, 1, 1, \ldots, 1}_{|A| \text{ times}}, \underbrace{0, 0, \ldots, 0}_{m-|A| \text{ times}}) \quad (16.23)$$

  so that $1_A(i) = 1$ if $i \leq |A|$, and $1_A(i) = 0$ otherwise.
- For any $f : 2^E \to \mathbb{R}$, $w = \mathbf{1}_A$, since $E_{|A|} = \{e_1, e_2, \ldots, e_{|A|}\} = A$:

$$\breve{f}(w)$$

From $\breve{f}$ back to f, even when f is not submodular

- From the continuous $\breve{f}$, we can recover $f(A)$ for any $A \subseteq V$.
- Take $w = \mathbf{1}_A$ for some $A \subseteq E$, so w is vertex of the hypercube.
- Order and rename the elements of E in decreasing order of w so that $w(e_1) \geq w(e_2) \geq w(e_3) \geq \cdots \geq w(e_m)$.
- This means

$$w = (w(e_1), w(e_2), \ldots, w(e_m)) = (\underbrace{1, 1, 1, \ldots, 1}_{|A| \text{ times}}, \underbrace{0, 0, \ldots, 0}_{m-|A| \text{ times}}) \quad (16.23)$$

  so that $1_A(i) = 1$ if $i \leq |A|$, and $1_A(i) = 0$ otherwise.
- For any $f : 2^E \to \mathbb{R}$, $w = \mathbf{1}_A$, since $E_{|A|} = \{e_1, e_2, \ldots, e_{|A|}\} = A$:

$$\breve{f}(w) = \sum_{i=1}^{m} \lambda_i f(E_i)$$

From $\breve{f}$ back to f, even when f is not submodular

- From the continuous $\breve{f}$, we can recover $f(A)$ for any $A \subseteq V$.
- Take $w = \mathbf{1}_A$ for some $A \subseteq E$, so w is vertex of the hypercube.
- Order and rename the elements of E in decreasing order of w so that $w(e_1) \geq w(e_2) \geq w(e_3) \geq \cdots \geq w(e_m)$.
- This means

$$w = (w(e_1), w(e_2), \ldots, w(e_m)) = (\underbrace{1, 1, 1, \ldots, 1}_{|A| \text{ times}}, \underbrace{0, 0, \ldots, 0}_{m-|A| \text{ times}}) \quad (16.23)$$

  so that $1_A(i) = 1$ if $i \leq |A|$, and $1_A(i) = 0$ otherwise.
- For any $f : 2^E \to \mathbb{R}$, $w = \mathbf{1}_A$, since $E_{|A|} = \{e_1, e_2, \ldots, e_{|A|}\} = A$:

$$\breve{f}(w) = \sum_{i=1}^{m} \lambda_i f(E_i) = w(e_m) f(E_m) + \sum_{i=1}^{m-1} (w(e_i) - w(e_{i+1}) f(E_i)$$

From $\breve{f}$ back to f, even when f is not submodular

- From the continuous $\breve{f}$, we can recover $f(A)$ for any $A \subseteq V$.
- Take $w = \mathbf{1}_A$ for some $A \subseteq E$, so w is vertex of the hypercube.
- Order and rename the elements of E in decreasing order of w so that $w(e_1) \geq w(e_2) \geq w(e_3) \geq \cdots \geq w(e_m)$.
- This means

$$w = (w(e_1), w(e_2), \ldots, w(e_m)) = (\underbrace{1, 1, 1, \ldots, 1}_{|A| \text{ times}}, \underbrace{0, 0, \ldots, 0}_{m-|A| \text{ times}}) \tag{16.23}$$

  so that $1_A(i) = 1$ if $i \leq |A|$, and $1_A(i) = 0$ otherwise.
- For any $f : 2^E \to \mathbb{R}$, $w = \mathbf{1}_A$, since $E_{|A|} = \{e_1, e_2, \ldots, e_{|A|}\} = A$:

$$\breve{f}(w) = \sum_{i=1}^{m} \lambda_i f(E_i) = w(e_m) f(E_m) + \sum_{i=1}^{m-1} (w(e_i) - w(e_{i+1}) f(E_i)$$

$$= \mathbf{1}_A(m) f(E_m) + \sum_{i=1}^{m-1} (\mathbf{1}_A(i) - \mathbf{1}_A(i+1)) f(E_i) \tag{16.24}$$

From $\breve{f}$ back to f, even when f is not submodular

- From the continuous $\breve{f}$, we can recover $f(A)$ for any $A \subseteq V$.
- Take $w = \mathbf{1}_A$ for some $A \subseteq E$, so w is vertex of the hypercube.
- Order and rename the elements of E in decreasing order of w so that $w(e_1) \geq w(e_2) \geq w(e_3) \geq \cdots \geq w(e_m)$.
- This means

$$w = (w(e_1), w(e_2), \ldots, w(e_m)) = (\underbrace{1, 1, 1, \ldots, 1}_{|A| \text{ times}}, \underbrace{0, 0, \ldots, 0}_{m-|A| \text{ times}}) \tag{16.23}$$

so that $1_A(i) = 1$ if $i \leq |A|$, and $1_A(i) = 0$ otherwise.

- For any $f : 2^E \to \mathbb{R}$, $w = \mathbf{1}_A$, since $E_{|A|} = \{e_1, e_2, \ldots, e_{|A|}\} = A$:

$$\begin{aligned}\breve{f}(w) &= \sum_{i=1}^{m} \lambda_i f(E_i) = w(e_m) f(E_m) + \sum_{i=1}^{m-1} (w(e_i) - w(e_{i+1}) f(E_i) \\ &= \mathbf{1}_A(m) f(E_m) + \sum_{i=1}^{m-1} (\mathbf{1}_A(i) - \mathbf{1}_A(i+1)) f(E_i) \qquad (16.24) \\ &= (\mathbf{1}_A(|A|) - \mathbf{1}_A(|A|+1)) f(E_{|A|}) = f(E_{|A|})\end{aligned}$$

From $\breve{f}$ back to f, even when f is not submodular

- From the continuous $\breve{f}$, we can recover $f(A)$ for any $A \subseteq V$.
- Take $w = \mathbf{1}_A$ for some $A \subseteq E$, so w is vertex of the hypercube.
- Order and rename the elements of E in decreasing order of w so that $w(e_1) \geq w(e_2) \geq w(e_3) \geq \cdots \geq w(e_m)$.
- This means

$$w = (w(e_1), w(e_2), \ldots, w(e_m)) = (\underbrace{1, 1, 1, \ldots, 1}_{|A| \text{ times}}, \underbrace{0, 0, \ldots, 0}_{m-|A| \text{ times}}) \tag{16.23}$$

so that $1_A(i) = 1$ if $i \leq |A|$, and $1_A(i) = 0$ otherwise.

- For any $f : 2^E \to \mathbb{R}$, $w = \mathbf{1}_A$, since $E_{|A|} = \{e_1, e_2, \ldots, e_{|A|}\} = A$:

$$\breve{f}(w) = \sum_{i=1}^{m} \lambda_i f(E_i) = w(e_m) f(E_m) + \sum_{i=1}^{m-1} (w(e_i) - w(e_{i+1}) f(E_i)$$

$$= \mathbf{1}_A(m) f(E_m) + \sum_{i=1}^{m-1} (\mathbf{1}_A(i) - \mathbf{1}_A(i+1)) f(E_i) \tag{16.24}$$

$$= (\mathbf{1}_A(|A|) - \mathbf{1}_A(|A|+1)) f(E_{|A|}) = f(E_{|A|}) = f(A) \tag{16.25}$$

Summary so far

- To summarize, with $\breve{f}(\mathbf{1}_A) = \sum_{i=1}^{m} \lambda_i f(E_i)$, we have

tight

$$\breve{f}(\mathbf{1}_A) = f(A), \quad \text{even if } f \text{ not submodular} \tag{16.26}$$

Summary so far

- To summarize, with $\breve{f}(\mathbf{1}_A) = \sum_{i=1}^m \lambda_i f(E_i)$, we have

$$\breve{f}(\mathbf{1}_A) = f(A), \tag{16.26}$$

- Thus, we can view $\breve{f} : [0,1]^E \to \mathbb{R}$ defined on the hypercube, with f defined as $\breve{f}$ evaluated on the hypercube extreme points (vertices).

Summary so far

- To summarize, with $\breve{f}(\mathbf{1}_A) = \sum_{i=1}^m \lambda_i f(E_i)$, we have

$$\breve{f}(\mathbf{1}_A) = f(A), \tag{16.26}$$

- Thus, we can view $\breve{f} : [0,1]^E \to \mathbb{R}$ defined on the hypercube, with f defined as $\breve{f}$ evaluated on the hypercube extreme points (vertices).
- ...and when f is submodular, we also have have

$$\breve{f}(\mathbf{1}_A) = \max\{\mathbf{1}_A{}^\intercal x : x \in B_f\} \tag{16.27}$$

and

$$\breve{f}(w) = \max\{w^\intercal x : x \in B_f\} = \sum_{i=1}^m \lambda_i f(E_i) \tag{16.28}$$

Summary so far

- To summarize, with $\breve{f}(\mathbf{1}_A) = \sum_{i=1}^m \lambda_i f(E_i)$, we have

$$\breve{f}(\mathbf{1}_A) = f(A), \tag{16.26}$$

- Thus, we can view $\breve{f} : [0,1]^E \to \mathbb{R}$ defined on the hypercube, with f defined as $\breve{f}$ evaluated on the hypercube extreme points (vertices).
- ...and when f is submodular, we also have have

$$\breve{f}(\mathbf{1}_A) = \max\{\mathbf{1}_A{}^\intercal x : x \in B_f\} \tag{16.27}$$

and

$$\breve{f}(w) = \max\{w^\intercal x : x \in B_f\} = \sum_{i=1}^m \lambda_i f(E_i) \tag{16.28}$$

- When considering $\breve{f} : [0,1]^E \to \mathbb{R}$, then any $w \in [0,1]^E$ is in positive orthant, and we have

$$\breve{f}(w) = \max\{w^\intercal x : x \in P_f\} \tag{16.29}$$

The Lovász extension of an arbitrary $f : 2^V \to \mathbb{R}$

- Thus, for any $f : 2^E \to \mathbb{R}$, even non-submodular f, we can define an extension, having $\breve{f}(\mathbf{1}_A) = f(A)$, $\forall A$, in this way where

$$\breve{f}(w) = \sum_{i=1}^{m} \lambda_i f(E_i) \tag{16.30}$$

with the $E_i = \{e_1, \dots, e_i\}$'s defined based on sorted descending order of w as in $w(e_1) \geq w(e_2) \geq \cdots \geq w(e_m)$, and where

$$\text{for } i \in \{1, \dots, m\}, \quad \lambda_i = \begin{cases} w(e_i) - w(e_{i+1}) & \text{if } i < m \\ w(e_m) & \text{if } i = m \end{cases} \tag{16.31}$$

so that $w = \sum_{i=1}^{m} \lambda_i \mathbf{1}_{E_i}$.

The Lovász extension of an arbitrary $f : 2^V \to \mathbb{R}$

- Thus, for any $f : 2^E \to \mathbb{R}$, even non-submodular f, we can define an extension, having $\breve{f}(\mathbf{1}_A) = f(A)$, $\forall A$, in this way where

$$\breve{f}(w) = \sum_{i=1}^{m} \lambda_i f(E_i) \tag{16.30}$$

with the $E_i = \{e_1, \dots, e_i\}$'s defined based on sorted descending order of w as in $w(e_1) \geq w(e_2) \geq \cdots \geq w(e_m)$, and where

$$\text{for } i \in \{1, \dots, m\}, \quad \lambda_i = \begin{cases} w(e_i) - w(e_{i+1}) & \text{if } i < m \\ w(e_m) & \text{if } i = m \end{cases} \tag{16.31}$$

so that $w = \sum_{i=1}^{m} \lambda_i \mathbf{1}_{E_i}$.

- $w = \sum_{i=1}^{m} \lambda_i \mathbf{1}_{E_i}$ is an interpolation of certain hypercube vertices.

The Lovász extension of an arbitrary $f : 2^V \to \mathbb{R}$

- Thus, for any $f : 2^E \to \mathbb{R}$, even non-submodular f, we can define an extension, having $\breve{f}(\mathbf{1}_A) = f(A)$, $\forall A$, in this way where

$$\breve{f}(w) = \sum_{i=1}^{m} \lambda_i f(E_i) \tag{16.30}$$

with the $E_i = \{e_1, \dots, e_i\}$'s defined based on sorted descending order of w as in $w(e_1) \geq w(e_2) \geq \cdots \geq w(e_m)$, and where

$$\text{for } i \in \{1, \dots, m\}, \quad \lambda_i = \begin{cases} w(e_i) - w(e_{i+1}) & \text{if } i < m \\ w(e_m) & \text{if } i = m \end{cases} \tag{16.31}$$

so that $w = \sum_{i=1}^{m} \lambda_i \mathbf{1}_{E_i}$.

- $w = \sum_{i=1}^{m} \lambda_i \mathbf{1}_{E_i}$ is an interpolation of certain hypercube vertices.
- $\breve{f}(w) = \sum_{i=1}^{m} \lambda_i f(E_i)$ is the associated interpolation of the values of f at sets corresponding to each hypercube vertex.

The Lovász extension of an arbitrary $f : 2^V \to \mathbb{R}$

- Thus, for any $f : 2^E \to \mathbb{R}$, even non-submodular f, we can define an extension, having $\breve{f}(\mathbf{1}_A) = f(A)$, $\forall A$, in this way where

$$\breve{f}(w) = \sum_{i=1}^{m} \lambda_i f(E_i) \tag{16.30}$$

with the $E_i = \{e_1, \dots, e_i\}$'s defined based on sorted descending order of w as in $w(e_1) \geq w(e_2) \geq \cdots \geq w(e_m)$, and where

$$\text{for } i \in \{1, \dots, m\}, \quad \lambda_i = \begin{cases} w(e_i) - w(e_{i+1}) & \text{if } i < m \\ w(e_m) & \text{if } i = m \end{cases} \tag{16.31}$$

so that $w = \sum_{i=1}^{m} \lambda_i \mathbf{1}_{E_i}$.

$w \in [0,1]^E \Rightarrow \lambda_i \geq 0 \quad \sum_i \lambda_i \leq 1$

if we set $\lambda_0 = \sum_{i=1}^{m} \lambda_i$

$\Rightarrow$ dist.

- $w = \sum_{i=1}^{m} \lambda_i \mathbf{1}_{E_i}$ is an interpolation of certain hypercube vertices.
- $\breve{f}(w) = \sum_{i=1}^{m} \lambda_i f(E_i)$ is the associated interpolation of the values of f at sets corresponding to each hypercube vertex.
- This extension is called the Lovász extension!

L.E.: Weighted gains or weighted functions, take your pick

- Again sorting E descending in w, the extension summarized:

$$\breve{f}(w) = \sum_{i=1}^{m} w(e_i) f(e_i | E_{i-1}) \qquad \text{weighted gains} \tag{16.32}$$

$$= \sum_{i=1}^{m} w(e_i)(f(E_i) - f(E_{i-1})) \tag{16.33}$$

$$= w(e_m) f(E_m) + \sum_{i=1}^{m-1} (w(e_i) - w(e_{i+1})) f(E_i) \tag{16.34}$$

$$= \sum_{i=1}^{m} \lambda_i f(E_i) \qquad \text{weighted functions} \tag{16.35}$$

L.E.: Weighted gains or weighted functions, take your pick

- Again sorting E descending in w, the extension summarized:

$$\breve{f}(w) = \sum_{i=1}^{m} w(e_i) f(e_i | E_{i-1}) \tag{16.32}$$

$$= \sum_{i=1}^{m} w(e_i)(f(E_i) - f(E_{i-1})) \tag{16.33}$$

$$= w(e_m) f(E_m) + \sum_{i=1}^{m-1} (w(e_i) - w(e_{i+1})) f(E_i) \tag{16.34}$$

$$= \sum_{i=1}^{m} \lambda_i f(E_i) \tag{16.35}$$

- So $\breve{f}(w)$ seen either as sum of weighted gain evaluations (Eqn. (16.32)), or as sum of weighted function evaluations (Eqn. (16.35)).

Summary: comparison of the two extension forms

- So if f is submodular, then we can write $\breve{f}(w) = \max(wx : x \in B_f)$ (which is clearly convex) in the form:

$$\breve{f}(w) = \max(wx : x \in B_f) = \sum_{i=1}^{m} \lambda_i f(E_i) \tag{16.36}$$

where $w = \sum_{i=1}^{m} \lambda_i \mathbf{1}_{E_i}$ and $E_i = \{e_1, \dots, e_i\}$ defined based on sorted descending order $w(e_1) \geq w(e_2) \geq \cdots \geq w(e_m)$.

Summary: comparison of the two extension forms

- So if f is submodular, then we can write $\breve{f}(w) = \max(wx : x \in B_f)$ (which is clearly convex) in the form:

$$\breve{f}(w) = \max(wx : x \in B_f) = \sum_{i=1}^{m} \lambda_i f(E_i) \tag{16.36}$$

where $w = \sum_{i=1}^{m} \lambda_i \mathbf{1}_{E_i}$ and $E_i = \{e_1, \dots, e_i\}$ defined based on sorted descending order $w(e_1) \geq w(e_2) \geq \cdots \geq w(e_m)$.

- On the other hand, for any f (even non-submodular), we can produce an extension $\breve{f}$ having the form

$$\breve{f}(w) = \sum_{i=1}^{m} \lambda_i f(E_i) \tag{16.37}$$

where $w = \sum_{i=1}^{m} \lambda_i \mathbf{1}_{E_i}$ and $E_i = \{e_1, \dots, e_i\}$ defined based on sorted descending order $w(e_1) \geq w(e_2) \geq \cdots \geq w(e_m)$.

Summary: comparison of the two extension forms

- So if f is submodular, then we can write $\breve{f}(w) = \max(wx : x \in B_f)$ (which is clearly convex) in the form:

$$\breve{f}(w) = \max(wx : x \in B_f) = \sum_{i=1}^{m} \lambda_i f(E_i) \tag{16.36}$$

  where $w = \sum_{i=1}^{m} \lambda_i \mathbf{1}_{E_i}$ and $E_i = \{e_1, \dots, e_i\}$ defined based on sorted descending order $w(e_1) \geq w(e_2) \geq \cdots \geq w(e_m)$.
- On the other hand, for any f (even non-submodular), we can produce an extension $\breve{f}$ having the form

$$\breve{f}(w) = \sum_{i=1}^{m} \lambda_i f(E_i) \tag{16.37}$$

  where $w = \sum_{i=1}^{m} \lambda_i \mathbf{1}_{E_i}$ and $E_i = \{e_1, \dots, e_i\}$ defined based on sorted descending order $w(e_1) \geq w(e_2) \geq \cdots \geq w(e_m)$.
- In both Eq. (16.36) and Eq. (16.37), we have $\breve{f}(\mathbf{1}_A) = f(A)$, $\forall A$, but Eq. (16.37), might not be convex for non-submodular f.

Summary: comparison of the two extension forms

- So if f is submodular, then we can write $\breve{f}(w) = \max(wx : x \in B_f)$ (which is clearly convex) in the form:

$$\breve{f}(w) = \max(wx : x \in B_f) = \sum_{i=1}^{m} \lambda_i f(E_i) \tag{16.36}$$

  where $w = \sum_{i=1}^{m} \lambda_i \mathbf{1}_{E_i}$ and $E_i = \{e_1, \dots, e_i\}$ defined based on sorted descending order $w(e_1) \geq w(e_2) \geq \cdots \geq w(e_m)$.
- On the other hand, for any f (even non-submodular), we can produce an extension $\breve{f}$ having the form

$$\breve{f}(w) = \sum_{i=1}^{m} \lambda_i f(E_i) \tag{16.37}$$

  where $w = \sum_{i=1}^{m} \lambda_i \mathbf{1}_{E_i}$ and $E_i = \{e_1, \dots, e_i\}$ defined based on sorted descending order $w(e_1) \geq w(e_2) \geq \cdots \geq w(e_m)$.
- In both Eq. (16.36) and Eq. (16.37), we have $\breve{f}(\mathbf{1}_A) = f(A)$, $\forall A$, but Eq. (16.37), might not be convex for non-submodular f.
- Submodularity is sufficient for convexity, but is it also necessary?

The Lovász extension of $f : 2^E \to \mathbb{R}$

- This continuous extension $\breve{f}$ of f, in any case (f being submodular or not), is typically called the Lovász extension of f (but also sometimes called the Choquet integral, or the Lovász-Edmonds extension since Edmonds showed many of the critical results).

The Lovász extension of $f : 2^E \to \mathbb{R}$

- This continuous extension $\breve{f}$ of f, in any case (f being submodular or not), is typically called the Lovász extension of f (but also sometimes called the Choquet integral, or the Lovász-Edmonds extension since Edmonds showed many of the critical results).
- Lovász showed that if a function $\breve{f}(w)$ defined as in Eqn. (16.30) is convex, then f must be submodular.

Lovász Extension, Submodularity and Convexity

Theorem 16.5.2

A function $f : 2^E \to \mathbb{R}$ is submodular iff its Lovász extension $\breve{f}$ of f is convex.

Proof.

- We've already seen that if f is submodular, its extension can be written via Eqn.(16.30) due to the greedy algorithm, and therefore is also equivalent to $\breve{f}(w) = \max\{wx : x \in P_f\}$, and thus is convex.

$\sum_i \lambda_i f(E_i)$

...

Lovász Extension, Submodularity and Convexity

Theorem 16.5.2

A function $f : 2^E \to \mathbb{R}$ is submodular iff its Lovász extension $\breve{f}$ of f is convex.

Proof.

- We've already seen that if f is submodular, its extension can be written via Eqn.(16.30) due to the greedy algorithm, and therefore is also equivalent to $\breve{f}(w) = \max\{wx : x \in P_f\}$, and thus is convex.
- Conversely, suppose the Lovász extension $\breve{f}(w) = \sum_i \lambda_i f(E_i)$ (of some function $f : 2^E \to \mathbb{R}$) is a convex function.

...

Lovász Extension, Submodularity and Convexity

Theorem 16.5.2

A function $f : 2^E \to \mathbb{R}$ is submodular iff its Lovász extension $\breve{f}$ of f is convex.

Proof.

- We've already seen that if f is submodular, its extension can be written via Eqn.(16.30) due to the greedy algorithm, and therefore is also equivalent to $\breve{f}(w) = \max\{wx : x \in P_f\}$, and thus is convex.
- Conversely, suppose the Lovász extension $\breve{f}(w) = \sum_i \lambda_i f(E_i)$ of some function $f : 2^E \to \mathbb{R}$ is a convex function.
- We note that, based on the extension definition, in particular the definition of the $\{\lambda_i\}_i$, we have that $\breve{f}(\alpha w) = \alpha \breve{f}(w)$ for any $\alpha \in \mathbb{R}_+$. I.e., f is a positively homogeneous convex function.

...

Lovász Extension, Submodularity and Convexity

...proof of Thm. 16.5.2 cont.

- Earlier, we saw that $\breve{f}(\mathbf{1}_A) = f(A)$ for all $A \subseteq E$.

...

Lovász Extension, Submodularity and Convexity

...proof of Thm. 16.5.2 cont.

- Earlier, we saw that $\breve{f}(\mathbf{1}_A) = f(A)$ for all $A \subseteq E$.
- Now, given $A, B \subseteq E$, we will show that

$$\breve{f}(\mathbf{1}_A + \mathbf{1}_B) = \breve{f}(\mathbf{1}_{A\cup B} + \mathbf{1}_{A\cap B}) \tag{16.38}$$

$$= f(A \cup B) + f(A \cap B). \tag{16.39}$$

...

Lovász Extension, Submodularity and Convexity

. . . proof of Thm. 16.5.2 cont.

- Earlier, we saw that $\breve{f}(\mathbf{1}_A) = f(A)$ for all $A \subseteq E$.
- Now, given $A, B \subseteq E$, we will show that

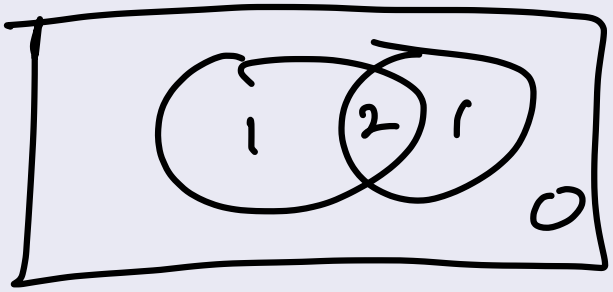
$$\breve{f}(\mathbf{1}_A + \mathbf{1}_B) = \breve{f}(\mathbf{1}_{A\cup B} + \mathbf{1}_{A\cap B}) \tag{16.38}$$

$$= f(A \cup B) + f(A \cap B). \tag{16.39}$$

- Let $C = A \cap B$, order E based on decreasing $w = \mathbf{1}_A + \mathbf{1}_B$ so that

$$w = (w(e_1), w(e_2), \dots, w(e_m)) \tag{16.40}$$

$$= (\underbrace{2, 2, \dots, 2}_{i\in C}, \underbrace{1, 1, \dots, 1}_{i\in A\triangle B}, \underbrace{0, 0, \dots, 0}_{i\in E\setminus(A\cup B)}) \tag{16.41}$$



. . .

Lovász Extension, Submodularity and Convexity

...proof of Thm. 16.5.2 cont.

- Earlier, we saw that $\breve{f}(\mathbf{1}_A) = f(A)$ for all $A \subseteq E$.
- Now, given $A, B \subseteq E$, we will show that

$$\breve{f}(\mathbf{1}_A + \mathbf{1}_B) = \breve{f}(\mathbf{1}_{A\cup B} + \mathbf{1}_{A\cap B}) \tag{16.38}$$

$$= f(A \cup B) + f(A \cap B). \tag{16.39}$$

- Let $C = A \cap B$, order E based on decreasing $w = \mathbf{1}_A + \mathbf{1}_B$ so that

$$w = (w(e_1), w(e_2), \dots, w(e_m)) \tag{16.40}$$

$$= (\underbrace{2, 2, \dots, 2}_{i\in C}, \underbrace{1, 1, \dots, 1}_{i\in A\triangle B}, \underbrace{0, 0, \dots, 0}_{i\in E\setminus(A\cup B)}) \tag{16.41}$$

- Then, considering $\breve{f}(w) = \sum_i \lambda_i f(E_i)$, we have $\lambda_{|C|} = 1$, $\lambda_{|A\cup B|} = 1$, and $\lambda_i = 0$ for $i \notin \{|C|, |A \cup B|\}$.

...

Lovász Extension, Submodularity and Convexity

...proof of Thm. 16.5.2 cont.

- Earlier, we saw that $\breve{f}(\mathbf{1}_A) = f(A)$ for all $A \subseteq E$.
- Now, given $A, B \subseteq E$, we will show that

$$\breve{f}(\mathbf{1}_A + \mathbf{1}_B) = \breve{f}(\mathbf{1}_{A\cup B} + \mathbf{1}_{A\cap B}) \tag{16.38}$$

$$= f(A \cup B) + f(A \cap B). \tag{16.39}$$

- Let $C = A \cap B$, order E based on decreasing $w = \mathbf{1}_A + \mathbf{1}_B$ so that

$$w = (w(e_1), w(e_2), \dots, w(e_m)) \tag{16.40}$$

$$= (\underbrace{2, 2, \dots, 2}_{i\in C}, \underbrace{1, 1, \dots, 1}_{i\in A\triangle B}, \underbrace{0, 0, \dots, 0}_{i\in E\setminus(A\cup B)}) \tag{16.41}$$

- Then, considering $\breve{f}(w) = \sum_i \lambda_i f(E_i)$, we have $\lambda_{|C|} = 1$, $\lambda_{|A\cup B|} = 1$, and $\lambda_i = 0$ for $i \notin \{|C|, |A \cup B|\}$.
- But then $E_{|C|} = A \cap B$ and $E_{|A\cup B|} = A \cup B$. Therefore, $\breve{f}(w) = \breve{f}(\mathbf{1}_A + \mathbf{1}_B) = f(A \cap B) + f(A \cup B)$.

...

Lovász Extension, Submodularity and Convexity

. . . proof of Thm. 16.5.2 cont.

- Also, since $\breve{f}$ is convex (by assumption) and positively homogeneous, we have for any $A, B \subseteq E$,

$$0.5[f(A \cap B) + f(A \cup B)] \tag{16.45}$$

□

Lovász Extension, Submodularity and Convexity

...proof of Thm. 16.5.2 cont.

- Also, since $\breve{f}$ is convex (by assumption) and positively homogeneous, we have for any $A, B \subseteq E$,

$$0.5[f(A \cap B) + f(A \cup B)] = 0.5[\breve{f}(\mathbf{1}_A + \mathbf{1}_B)] \tag{16.42}$$

(16.45)

□

Lovász Extension, Submodularity and Convexity

...proof of Thm. 16.5.2 cont.

- Also, since $\breve{f}$ is convex (by assumption) and positively homogeneous, we have for any $A, B \subseteq E$,

$$0.5[f(A \cap B) + f(A \cup B)] = 0.5[\breve{f}(\mathbf{1}_A + \mathbf{1}_B)] \tag{16.42}$$

$$= \breve{f}(0.5\mathbf{1}_A + 0.5\mathbf{1}_B) \tag{16.43}$$

$$\tag{16.45}$$

□

Lovász Extension, Submodularity and Convexity

. . . proof of Thm. 16.5.2 cont.

- Also, since $\breve{f}$ is convex (by assumption) and positively homogeneous, we have for any $A, B \subseteq E$,

$$0.5[f(A \cap B) + f(A \cup B)] = 0.5[\breve{f}(\mathbf{1}_A + \mathbf{1}_B)] \quad (16.42)$$

$$= \breve{f}(0.5\mathbf{1}_A + 0.5\mathbf{1}_B) \quad (16.43)$$

$$\leq 0.5\breve{f}(\mathbf{1}_A) + 0.5\breve{f}(\mathbf{1}_B) \quad (16.44)$$

$$(16.45)$$

□

Lovász Extension, Submodularity and Convexity

...proof of Thm. 16.5.2 cont.

- Also, since $\breve{f}$ is convex (by assumption) and positively homogeneous, we have for any $A, B \subseteq E$,

$$0.5[f(A \cap B) + f(A \cup B)] = 0.5[\breve{f}(\mathbf{1}_A + \mathbf{1}_B)] \quad (16.42)$$
$$= \breve{f}(0.5\mathbf{1}_A + 0.5\mathbf{1}_B) \quad (16.43)$$
$$\leq 0.5\breve{f}(\mathbf{1}_A) + 0.5\breve{f}(\mathbf{1}_B) \quad (16.44)$$
$$= 0.5(f(A) + f(B)) \quad (16.45)$$

□

Lovász Extension, Submodularity and Convexity

...proof of Thm. 16.5.2 cont.

- Also, since $\breve{f}$ is convex (by assumption) and positively homogeneous, we have for any $A, B \subseteq E$,

$$0.5[f(A \cap B) + f(A \cup B)] = 0.5[\breve{f}(\mathbf{1}_A + \mathbf{1}_B)] \tag{16.42}$$
$$= \breve{f}(0.5\mathbf{1}_A + 0.5\mathbf{1}_B) \tag{16.43}$$
$$\leq 0.5\breve{f}(\mathbf{1}_A) + 0.5\breve{f}(\mathbf{1}_B) \tag{16.44}$$
$$= 0.5(f(A) + f(B)) \tag{16.45}$$

- Thus, we have shown that for any $A, B \subseteq E$,

$$f(A \cup B) + f(A \cap B) \leq f(A) + f(B) \tag{16.46}$$

so f must be submodular.

□

Lovász ext. vs. the concave closure of submodular function

- The above theorem showed that the Lovász extension is convex iff f is submodular.

Lovász ext. vs. the concave closure of submodular function

- The above theorem showed that the Lovász extension is convex iff f is submodular.
- Our next theorem shows that the Lovász extension coincides precisely with the convex closure iff f is submodular.

Lovász ext. vs. the concave closure of submodular function

- The above theorem showed that the Lovász extension is convex iff f is submodular.
- Our next theorem shows that the Lovász extension coincides precisely with the convex closure iff f is submodular.
- I.e., not only is the Lovász extension convex for f submodular and tight at each $\mathbf{1}_A$, it is the actual convex closure when f.

Lovász ext. vs. the concave closure of submodular function

- The above theorem showed that the Lovász extension is convex iff f is submodular.
- Our next theorem shows that the Lovász extension coincides precisely with the convex closure iff f is submodular.
- I.e., not only is the Lovász extension convex for f submodular and tight at each $\mathbf{1}_A$, it is the actual convex closure when f.
- Hence, convex closure (which normally looks daunting) is easy to evaluate (the greedy algorithm) when f is submodular and takes this particular form iff f is submodular.

Lovász ext. vs. the concave closure of submodular function

Theorem 16.5.3

Let $\breve{f}(w) = \max(wy : y \in B_f) = \sum_{i=1}^{m} \lambda_i f(E_i)$ *be the Lovász extension and* $\check{f}(x) = \min_{p \in \triangle^n(x)} E_{S \sim p}[f(S)]$ *be the convex closure. Then* $\breve{f}$ *and* $\check{f}$ *coincide iff* f *is submodular, i.e.,* $\breve{f}(w) = \check{f}(w), \forall w \in [0, 1]$.

Proof.

- Assume f is submodular.

...

Lovász ext. vs. the concave closure of submodular function

Theorem 16.5.3

Let $\breve{f}(w) = \max(wy : y \in B_f) = \sum_{i=1}^{m} \lambda_i f(E_i)$ *be the Lovász extension and* $\check{f}(x) = \min_{p \in \triangle^n(x)} E_{S \sim p}[f(S)]$ *be the convex closure. Then* $\breve{f}$ *and* $\check{f}$ *coincide iff* f *is submodular, i.e.,* $\breve{f}(w) = \check{f}(w), \forall w \in [0,1]$.

Proof.

- Assume f is submodular.
- Given x, let p^x be an achieving argmin in $\check{f}(x)$ that also maximizes $\sum_S p^x_S |S|^2$.

...

Lovász ext. vs. the concave closure of submodular function

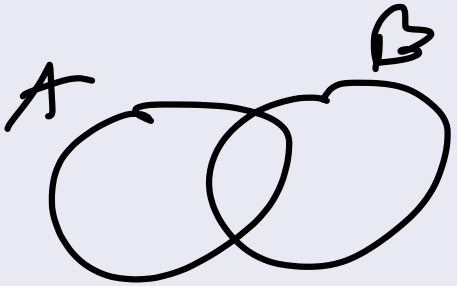
Theorem 16.5.3

Let $\breve{f}(w) = \max(wy : y \in B_f) = \sum_{i=1}^{m} \lambda_i f(E_i)$ *be the Lovász extension and* $\check{f}(x) = \min_{p \in \triangle^n(x)} E_{S \sim p}[f(S)]$ *be the convex closure. Then* $\breve{f}$ *and* $\check{f}$ *coincide iff* f *is submodular, i.e.,* $\breve{f}(w) = \check{f}(w), \forall w \in [0,1]$.

Proof.

- Assume f is submodular.
- Given x, let p^x be an achieving argmin in $\check{f}(x)$ that also maximizes $\sum_S p_S^x |S|^2$.
- Suppose $\exists A, B \subseteq V$ that are crossing (i.e., $A \not\subseteq B$, $B \not\subseteq A$) with positive p_A^x, p_B^x. W.l.o.g., $p_A^x \geq p_B^x > 0$.

Crossing



$\therefore$

$|A \setminus B| > 0$

$|B \setminus A| > 0$

...

Lovász ext. vs. the concave closure of submodular function

Theorem 16.5.3

Let $\breve{f}(w) = \max(wy : y \in B_f) = \sum_{i=1}^{m} \lambda_i f(E_i)$ *be the Lovász extension and* $\check{f}(x) = \min_{p \in \triangle^n(x)} E_{S \sim p}[f(S)]$ *be the convex closure. Then* $\breve{f}$ *and* $\check{f}$ *coincide iff* f *is submodular, i.e.,* $\breve{f}(w) = \check{f}(w), \forall w \in [0,1]$.

Proof.

- Assume f is submodular.
- Given x, let p^x be an achieving argmin in $\check{f}(x)$ that also maximizes $\sum_S p^x_S |S|^2$.
- Suppose $\exists A, B \subseteq V$ that are crossing (i.e., $A \not\subseteq B$, $B \not\subseteq A$) with positive p^x_A, p^x_B. W.l.o.g., $p^x_A \geq p^x_B > 0$.
- Then we may update p^x, keeping it a distribution, as follows:

$$\bar{p}^x_A \leftarrow p^x_A - p^x_B \geq 0 \qquad \bar{p}^x_B \leftarrow p^x_B - p^x_B = 0 \tag{16.47}$$

$$\bar{p}^x_{A \cup B} \leftarrow p^x_{A \cup B} + p^x_B \geq 0 \qquad \bar{p}^x_{A \cap B} \leftarrow p^x_{A \cap B} + p^x_B \geq 0 \tag{16.48}$$

and by submodularity, this does not increase $\sum_S p^x_S f(S)$.

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

...

Lovász ext. vs. the concave closure of submodular function

. . . proof cont.

- This does increase $\sum_S p_S^x |S|^2$ however since $|S|^2$ is supermodular, i.e.:

$$|A \cup B|^2 + |A \cap B|^2 = (|A| + |B \setminus A|)^2 + (|B| - |B \setminus A|)^2 \quad (16.49)$$

$$= |A|^2 + |B|^2 + 2|B \setminus A|(|A| - |B| + |B \setminus A|) \quad (16.50)$$

$$= |A|^2 + |B|^2 + 2\underbrace{|B \setminus A||A \setminus B|}_{\text{since } A \text{ crosses } B \; > 0} \quad (16.51)$$

$$> |A|^2 + |B|^2 \quad (16.52)$$

□

Lovász ext. vs. the concave closure of submodular function

. . . proof cont.

- This does increase $\sum_S p_S^x |S|^2$ however since $|S|^2$ is supermodular, i.e.:

$$|A \cup B|^2 + |A \cap B|^2 = (|A| + |B \setminus A|)^2 + (|B| - |B \setminus A|)^2 \quad (16.49)$$

$$= |A|^2 + |B|^2 + 2|B \setminus A|(|A| - |B| + |B \setminus A|) \quad (16.50)$$

$$= |A|^2 + |B|^2 + 2|B \setminus A||A \setminus B| \quad (16.51)$$

$$> |A|^2 + |B|^2 \quad (16.52)$$

- Contradiction! Hence, there can be no crossing sets A, B and we must have, for any A, B with $p_A^x > 0$ and $p_B^x > 0$ either $A \subset B$ or $B \subset A$.

□

Lovász ext. vs. the concave closure of submodular function

. . . proof cont.

- This does increase $\sum_S p_S^x |S|^2$ however since $|S|^2$ is supermodular, i.e.:

$$|A \cup B|^2 + |A \cap B|^2 = (|A| + |B \setminus A|)^2 + (|B| - |B \setminus A|)^2 \quad (16.49)$$
$$= |A|^2 + |B|^2 + 2|B \setminus A|(|A| - |B| + |B \setminus A|) \quad (16.50)$$
$$= |A|^2 + |B|^2 + 2|B \setminus A||A \setminus B| \quad (16.51)$$
$$> |A|^2 + |B|^2 \quad (16.52)$$

- Contradiction! Hence, there can be no crossing sets A, B and we must have, for any A, B with $p_A^x > 0$ and $p_B^x > 0$ either $A \subset B$ or $B \subset A$.
- Hence, the sets $\{A \subseteq V : p_A^x > 0\}$ form a chain and can be at most as large as size $n = |V|$.

$$\emptyset = E_0 \subset E_1 \subset E_2 \subset \cdots \subset E$$

□

Lovász ext. vs. the concave closure of submodular function

. . . proof cont.

- This does increase $\sum_S p_S^x |S|^2$ however since $|S|^2$ is supermodular, i.e.:

$$|A \cup B|^2 + |A \cap B|^2 = (|A| + |B \setminus A|)^2 + (|B| - |B \setminus A|)^2 \quad (16.49)$$

$$= |A|^2 + |B|^2 + 2|B \setminus A|(|A| - |B| + |B \setminus A|) \quad (16.50)$$

$$= |A|^2 + |B|^2 + 2|B \setminus A||A \setminus B| \quad (16.51)$$

$$> |A|^2 + |B|^2 \quad (16.52)$$

- Contradiction! Hence, there can be no crossing sets A, B and we must have, for any A, B with $p_A^x > 0$ and $p_B^x > 0$ either $A \subset B$ or $B \subset A$.
- Hence, the sets $\{A \subseteq V : p_A^x > 0\}$ form a chain and can be at most as large as size $n = |V|$.
- The only feasible chain is the same chain that defines the Lovász extension $\breve{f}(x)$, namely $\emptyset = E_0 \subseteq E_1 \subseteq E_2 \subset \dots$ where $E_i = \{e_1, e_2, \dots, e_i\}$ and e_i is ordered as $x(e_1) \geq x(e_2) \geq \dots \geq x(e_n)$.

□

Lovász ext. vs. the concave closure of submodular function

. . . proof cont.

- Next, assume f is not submodular. We must show that the Lovász extension $\breve{f}(x)$ and the concave closure $\check{f}(x)$ need not coincide.

Lovász ext. vs. the concave closure of submodular function

. . . proof cont.

- Next, assume f is not submodular. We must show that the Lovász extension $\breve{f}(x)$ and the concave closure $\check{f}(x)$ need not coincide.
- Since f is not submodular, $\exists S$ and $i, j \notin S$ such that $f(S) + f(S + i + j) > f(S + i) + f(S + j)$, a strict violation of submodularity.

Lovász ext. vs. the concave closure of submodular function

. . . proof cont.

- Next, assume f is not submodular. We must show that the Lovász extension $\breve{f}(x)$ and the concave closure $\check{f}(x)$ need not coincide.
- Since f is not submodular, $\exists S$ and $i, j \notin S$ such that $f(S) + f(S + i + j) > f(S + i) + f(S + j)$, a strict violation of submodularity.
- Consider $x = \mathbf{1}_S + \frac{1}{2}\mathbf{1}_{\{i,j\}}$. $\in [0,1]^E$

Lovász ext. vs. the concave closure of submodular function

. . . proof cont.

- Next, assume f is not submodular. We must show that the Lovász extension $\breve{f}(x)$ and the concave closure $\check{f}(x)$ need not coincide.
- Since f is not submodular, $\exists S$ and $i, j \notin S$ such that $f(S) + f(S+i+j) > f(S+i) + f(S+j)$, a strict violation of submodularity.
- Consider $x = \mathbf{1}_S + \frac{1}{2}\mathbf{1}_{\{i,j\}}$.

$$\underbrace{1\,1\,1\,1\,1\,1\,1\,1}_{S}\ \underbrace{1/2\ \ 1/2}_{i,j}\ \underbrace{0\,0\,0\,0\,0\,0}_{\text{rest}}$$

- Then L.E. has $\breve{f}(x) = \frac{1}{2}f(S) + \frac{1}{2}f(S+i+j)$ and this p is feasible for $\check{f}(x)$ with $p_S = 1/2$ and $p_{S+i+j} = 1/2$. $\in \mathcal{D}(x)$

Lovász ext. vs. the concave closure of submodular function

. . . proof cont.

- Next, assume f is not submodular. We must show that the Lovász extension $\breve{f}(x)$ and the concave closure $\check{f}(x)$ need not coincide.
- Since f is not submodular, $\exists S$ and $i, j \notin S$ such that $f(S) + f(S + i + j) > f(S + i) + f(S + j)$, a strict violation of submodularity.
- Consider $x = \mathbf{1}_S + \frac{1}{2}\mathbf{1}_{\{i,j\}}$.
- Then L.E. has $\breve{f}(x) = \frac{1}{2}f(S) + \frac{1}{2}f(S + i + j)$ and this p is feasible for $\check{f}(x)$ with $p_S = 1/2$ and $p_{S+i+j} = 1/2$.
- An alternate feasible distribution for $\check{f}(x)$ in the convex closure is $\bar{p}_{S+i} = \bar{p}_{S+j} = 1/2$.

Lovász ext. vs. the concave closure of submodular function

. . . proof cont.

- Next, assume f is not submodular. We must show that the Lovász extension $\breve{f}(x)$ and the concave closure $\check{f}(x)$ need not coincide.
- Since f is not submodular, $\exists S$ and $i, j \notin S$ such that $f(S) + f(S + i + j) > f(S + i) + f(S + j)$, a strict violation of submodularity.
- Consider $x = \mathbf{1}_S + \frac{1}{2}\mathbf{1}_{\{i,j\}}$.
- Then L.E. has $\breve{f}(x) = \frac{1}{2}f(S) + \frac{1}{2}f(S + i + j)$ and this p is feasible for $\check{f}(x)$ with $p_S = 1/2$ and $p_{S+i+j} = 1/2$.
- An alternate feasible distribution for $\check{f}(x)$ in the convex closure is $\bar{p}_{S+i} = \bar{p}_{S+j} = 1/2$.
- This gives

$$\check{f}(x) \leq \frac{1}{2}[f(S + i) + f(S + j)] < \breve{f}(x) \tag{16.53}$$

meaning $\check{f}(x) \neq \breve{f}(x)$. □