Submodular Functions, Optimization, and Applications to Machine Learning

— Spring Quarter, Lecture 15 —

http://j.ee.washington.edu/~bilmes/classes/ee596b_spring_2014/

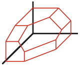
Prof. Jeff Bilmes

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

May 19th, 2014

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

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



Cumulative Outstanding Reading

- Good references for today: Schrijver-2003, Oxley-1992/2011, Welsh-1973, Goemans-2010, Cunningham-1984, Edmonds-1969.
- Read Tom McCormick's overview paper on SFM http://people.commerce.ubc.ca/faculty/mccormick/sfmchap8a.pdf
- Read chapters 1 - 4 from Fujishige book.
- Matroid properties http://www-math.mit.edu/~goemans/18433S09/matroid-notes.pdf
- Read lecture 14 slides on lattice theory at our web page (http://j.ee.washington.edu/~bilmes/classes/ee596b_spring_2014/)

Announcements, Assignments, and Reminders

- Weekly Office Hours: Wednesdays, 5:00-5:50, or by skype or google hangout (email me).

Class Road Map - IT-I

- L1 (3/31): Motivation, Applications, & Basic Definitions
- L2: (4/2): Applications, Basic Definitions, Properties
- L3: More examples and properties (e.g., closure properties), and examples, spanning trees
- L4: proofs of equivalent definitions, independence, start matroids
- L5: matroids, basic definitions and examples
- L6: More on matroids, System of Distinct Reps, Transversals, Transversal Matroid, Matroid and representation
- L7: Dual Matroids, other matroid properties, Combinatorial Geometries
- L8: Combinatorial Geometries, matroids and greedy, Polyhedra, Matroid Polytopes,
- L9: From Matroid Polytopes to Polymatroids.
- L10: Polymatroids and Submodularity
- L11: More properties of polymatroids, SFM special cases
- L12: polymatroid properties, extreme points polymatroids,
- L13: sat, dep, supp, exchange capacity, examples
- L14: Lattice theory: partially ordered sets; lattices; distributive, modular, submodular, and boolean lattices; ideals and join irreducibles.
- L15: Supp, Base polytope, exchange capacity, minimum norm point algorithm and the lattice of minimizers of a submodular function, Lovasz extension
- L16:
- L17:
- L18:
- L19:
- L20:

Finals Week: June 9th-13th, 2014.

Summary of Concepts

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

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

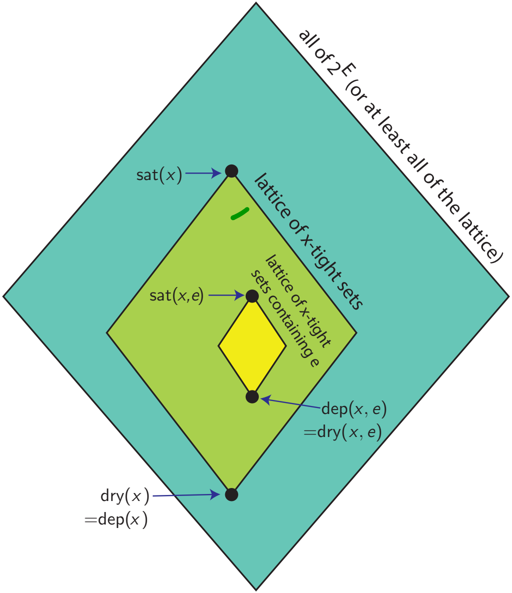
- x-tight sets: For $x \in P_f$, $\mathcal{D}(x) = \{A \subseteq E : x(A) = f(A)\}$.
- Polymatroid closure/maximal x-tight set: For $x \in P_f$, $\text{sat}(x) = \cup\{A : A \in \mathcal{D}(x)\} = \{e : e \in E, \forall\alpha > 0, x + \alpha \mathbf{1}_e \notin P_f\}$.
- Saturation capacity: for $x \in P_f$, $0 \le \hat{c}(x; e) = \min \{f(A) - x(A) | \forall A \ni e\} = \max \{\alpha : \alpha \in \mathbb{R}, x + \alpha \mathbf{1}_e \in P_f\}$
- Recall: $\text{sat}(x) = \{e : \hat{c}(x; e) = 0\}$ and $E \setminus \text{sat}(x) = \{e : \hat{c}(x; e) > 0\}$.
- e-containing x-tight sets: For $x \in P_f$, $\mathcal{D}(x, e) = \{A : e \in A \subseteq E, x(A) = f(A)\} \subseteq \mathcal{D}(x)$.
- Minimal e-containing x-tight set/polymatroidal fundamental circuit/: For $x \in P_f$,

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

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

dep and sat in a lattice

- Given some $x \in P_f$,
- The picture on the right summarizes the relationships between the lattices and sublattices.
- Note, $\bigcap_e \mathrm{dep}(x, e) = \mathrm{dep}(x)$.



Support of vector

- The support of a vector $x \in P_f$ is defined as the elements with non-zero entries.

Support of vector

- The support of a vector $x \in P_f$ is defined as the elements with non-zero entries.
- That is

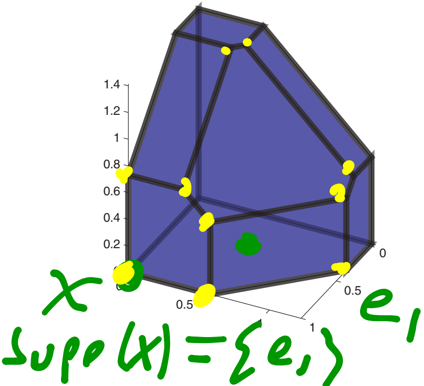
$$\text{supp}(x) = \{e \in E : x(e) \neq 0\} \tag{15.1}$$

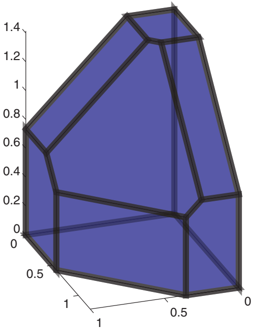
Support of vector

- The support of a vector $x \in P_f$ is defined as the elements with non-zero entries.
- That is

$$\text{supp}(x) = \{e \in E : x(e) \neq 0\} \tag{15.1}$$

- Example





Tightness of supp at polymatroidal extreme point

- Now, $\text{sat}(x)$ is tight, and corresponds to the largest member of the distributive lattice $\mathcal{D}(x) = \{A : x(A) = f(A)\}$ of tight sets.

Tightness of supp at polymatroidal extreme point

- Now, $\text{sat}(x)$ is tight, and corresponds to the largest member of the distributive lattice $\mathcal{D}(x) = \{A : x(A) = f(A)\}$ of tight sets.
- $\text{supp}(x)$ is not necessarily tight for an arbitrary x.

Tightness of supp at polymatroidal extreme point

- Now, $\mathrm{sat}(x)$ is tight, and corresponds to the largest member of the distributive lattice $\mathcal{D}(x) = \{A : x(A) = f(A)\}$ of tight sets.
- $\mathrm{supp}(x)$ is not necessarily tight for an arbitrary x.
- When x is an extremal point, however, $\mathrm{supp}(x)$ is tight, meaning $x(\mathrm{supp}(x)) = f(\mathrm{supp}(x))$. Why?

Tightness of supp at polymatroidal extreme point

- Now, $\text{sat}(x)$ is tight, and corresponds to the largest member of the distributive lattice $\mathcal{D}(x) = \{A : x(A) = f(A)\}$ of tight sets.
- $\text{supp}(x)$ is not necessarily tight for an arbitrary x.
- When x is an extremal point, however, $\text{supp}(x)$ is tight, meaning $x(\text{supp}(x)) = f(\text{supp}(x))$. Why?
  1. Extremal points are defined as a system of equalities of the form $x(E_i) = f(E_i)$ for $1 \leq i \leq k \leq |E|$, for some k, as we saw earlier in class. Hence, any $e_i \in \text{supp}(x)$ has $x(e_i) = f(e_i|E_{i-1}) > 0$.

Tightness of supp at polymatroidal extreme point

- Now, $\mathrm{sat}(x)$ is tight, and corresponds to the largest member of the distributive lattice $\mathcal{D}(x) = \{A : x(A) = f(A)\}$ of tight sets.
- $\mathrm{supp}(x)$ is not necessarily tight for an arbitrary x.
- When x is an extremal point, however, $\mathrm{supp}(x)$ is tight, meaning $x(\mathrm{supp}(x)) = f(\mathrm{supp}(x))$. Why?
  1. Extremal points are defined as a system of equalities of the form $x(E_i) = f(E_i)$ for $1 \le i \le k \le |E|$, for some k, as we saw earlier in class. Hence, any $e_i \in \mathrm{supp}(x)$ has $x(e_i) = f(e_i|E_{i-1}) > 0$.
  2. Now, for $1 \le i \le k$, if $e_i \notin \mathrm{supp}(x)$, $x(E_k) = x(E_k - e_i)$

Tightness of supp at polymatroidal extreme point

- Now, $\text{sat}(x)$ is tight, and corresponds to the largest member of the distributive lattice $\mathcal{D}(x) = \{A : x(A) = f(A)\}$ of tight sets.
- $\text{supp}(x)$ is not necessarily tight for an arbitrary x.
- When x is an extremal point, however, $\text{supp}(x)$ is tight, meaning $x(\text{supp}(x)) = f(\text{supp}(x))$. Why?
  1. Extremal points are defined as a system of equalities of the form $x(E_i) = f(E_i)$ for $1 \leq i \leq k \leq |E|$, for some k, as we saw earlier in class. Hence, any $e_i \in \text{supp}(x)$ has $x(e_i) = f(e_i|E_{i-1}) > 0$.
  2. Now, for $1 \leq i \leq k$, if $e_i \notin \text{supp}(x)$, $\boxed{x(E_k) = x(E_k - e_i)}$
  3. Also, for $1 \leq i \leq k$, if $e_i \notin \text{supp}(x)$, then $x(e_i) = 0 = f(e_i|E_{i-1}) \geq f(e_i|E_k - e_i) = f(E_k|E_k - e_i) \geq 0$ since monotone submodular, hence we have $\boxed{f(E_k) = f(E_k - e_i)}$.

Tightness of supp at polymatroidal extreme point

- Now, $\text{sat}(x)$ is tight, and corresponds to the largest member of the distributive lattice $\mathcal{D}(x) = \{A : x(A) = f(A)\}$ of tight sets.
- $\text{supp}(x)$ is not necessarily tight for an arbitrary x.
- When x is an extremal point, however, $\text{supp}(x)$ is tight, meaning $x(\text{supp}(x)) = f(\text{supp}(x))$. Why?
  1. Extremal points are defined as a system of equalities of the form $x(E_i) = f(E_i)$ for $1 \leq i \leq k \leq |E|$, for some k, as we saw earlier in class. Hence, any $e_i \in \text{supp}(x)$ has $x(e_i) = f(e_i|E_{i-1}) > 0$.
  2. Now, for $1 \leq i \leq k$, if $e_i \notin \text{supp}(x)$, $x(E_k) = x(E_k - e_i)$
  3. Also, for $1 \leq i \leq k$, if $e_i \notin \text{supp}(x)$, then $x(e_i) = 0 = f(e_i|E_{i-1}) \geq f(e_i|E_k - e_i) = f(E_k|E_k - e_i) \geq 0$ since monotone submodular, hence we have $f(E_k) = f(E_k - e_i)$.
  4. Thus, $x(E_k - e_i) = f(E_k - e_i)$ and $E_k - e_i$ is also tight.

Tightness of supp at polymatroidal extreme point

- Now, $\text{sat}(x)$ is tight, and corresponds to the largest member of the distributive lattice $\mathcal{D}(x) = \{A : x(A) = f(A)\}$ of tight sets.
- $\text{supp}(x)$ is not necessarily tight for an arbitrary x.
- When x is an extremal point, however, $\text{supp}(x)$ is tight, meaning $x(\text{supp}(x)) = f(\text{supp}(x))$. Why?
  1. Extremal points are defined as a system of equalities of the form $x(E_i) = f(E_i)$ for $1 \leq i \leq k \leq |E|$, for some k, as we saw earlier in class. Hence, any $e_i \in \text{supp}(x)$ has $x(e_i) = f(e_i|E_{i-1}) > 0$.
  2. Now, for $1 \leq i \leq k$, if $e_i \notin \text{supp}(x)$, $x(E_k) = x(E_k - e_i)$
  3. Also, for $1 \leq i \leq k$, if $e_i \notin \text{supp}(x)$, then $x(e_i) = 0 = f(e_i|E_{i-1}) \geq f(e_i|E_k - e_i) = f(E_k|E_k - e_i) \geq 0$ since monotone submodular, hence we have $f(E_k) = f(E_k - e_i)$.
  4. Thus, $x(E_k - e_i) = f(E_k - e_i)$ and $E_k - e_i$ is also tight.
  5. We can keep removing elements $\notin \text{supp}(x)$ and we're left with $f(E_k \cap \text{supp}(x)) = x(E_k \cap \text{supp}(x))$ for any k.

Tightness of supp at polymatroidal extreme point

- Now, $\text{sat}(x)$ is tight, and corresponds to the largest member of the distributive lattice $\mathcal{D}(x) = \{A : x(A) = f(A)\}$ of tight sets.
- $\text{supp}(x)$ is not necessarily tight for an arbitrary x.
- When x is an extremal point, however, $\text{supp}(x)$ is tight, meaning $x(\text{supp}(x)) = f(\text{supp}(x))$. Why?
  1. Extremal points are defined as a system of equalities of the form $x(E_i) = f(E_i)$ for $1 \leq i \leq k \leq |E|$, for some k, as we saw earlier in class. Hence, any $e_i \in \text{supp}(x)$ has $x(e_i) = f(e_i|E_{i-1}) > 0$.
  2. Now, for $1 \leq i \leq k$, if $e_i \notin \text{supp}(x)$, $x(E_k) = x(E_k - e_i)$
  3. Also, for $1 \leq i \leq k$, if $e_i \notin \text{supp}(x)$, then $x(e_i) = 0 = f(e_i|E_{i-1}) \geq f(e_i|E_k - e_i) = f(E_k|E_k - e_i) \geq 0$ since monotone submodular, hence we have $f(E_k) = f(E_k - e_i)$.
  4. Thus, $x(E_k - e_i) = f(E_k - e_i)$ and $E_k - e_i$ is also tight.
  5. We can keep removing elements $\notin \text{supp}(x)$ and we're left with $f(E_k \cap \text{supp}(x)) = x(E_k \cap \text{supp}(x))$ for any k.
  6. Hence $\text{supp}(x)$ is tight when x is extremal.

Tightness of supp at polymatroidal extreme point

- Now, $\text{sat}(x)$ is tight, and corresponds to the largest member of the distributive lattice $\mathcal{D}(x) = \{A : x(A) = f(A)\}$ of tight sets.
- $\text{supp}(x)$ is not necessarily tight for an arbitrary x.
- When x is an extremal point, however, $\text{supp}(x)$ is tight, meaning $x(\text{supp}(x)) = f(\text{supp}(x))$. Why?
  1. Extremal points are defined as a system of equalities of the form $x(E_i) = f(E_i)$ for $1 \leq i \leq k \leq |E|$, for some k, as we saw earlier in class. Hence, any $e_i \in \text{supp}(x)$ has $x(e_i) = f(e_i|E_{i-1}) > 0$.
  2. Now, for $1 \leq i \leq k$, if $e_i \notin \text{supp}(x)$, $x(E_k) = x(E_k - e_i)$
  3. Also, for $1 \leq i \leq k$, if $e_i \notin \text{supp}(x)$, then $x(e_i) = 0 = f(e_i|E_{i-1}) \geq f(e_i|E_k - e_i) = f(E_k|E_k - e_i) \geq 0$ since monotone submodular, hence we have $f(E_k) = f(E_k - e_i)$.
  4. Thus, $x(E_k - e_i) = f(E_k - e_i)$ and $E_k - e_i$ is also tight.
  5. We can keep removing elements $\notin \text{supp}(x)$ and we're left with $f(E_k \cap \text{supp}(x)) = x(E_k \cap \text{supp}(x))$ for any k.
  6. Hence $\text{supp}(x)$ is tight when x is extremal.
- Since $\text{supp}(x)$ is tight, we immediately have that $\text{sat}(x) \supseteq \text{supp}(x)$.

supp vs. sat equality

- For $x \in P_f$, with x extremal, is $\mathrm{supp}(x) = \mathrm{sat}(x)$?

supp vs. sat equality

- For $x \in P_f$, with x extremal, is $\text{supp}(x) = \text{sat}(x)$?
- Consider an example case where disjoint $X, Y \subseteq E$, we have $f(X) = f(Y) = f(X \cup Y)$ (meaning "perfect dependence" or full redundancy, so gains are not strictly positive), $f(Y|X) = 0$.

supp vs. sat equality

- For $x \in P_f$, with x extremal, is $\text{supp}(x) = \text{sat}(x)$?
- Consider an example case where disjoint $X, Y \subseteq E$, we have $f(X) = f(Y) = f(X \cup Y)$ (meaning "perfect dependence" or full redundancy, so gains are not strictly positive), $f(Y|X) = 0$.
- Suppose $x \in P_f$ has $x(X) > 0$ but $x(V \setminus X) = 0$ and so $x(Y) = 0$.

supp vs. sat equality

- For $x \in P_f$, with x extremal, is $\mathrm{supp}(x) = \mathrm{sat}(x)$?
- Consider an example case where disjoint $X, Y \subseteq E$, we have $f(X) = f(Y) = f(X \cup Y)$ (meaning "perfect dependence" or full redundancy, so gains are not strictly positive), $f(Y|X) = 0$.
- Suppose $x \in P_f$ has $x(X) > 0$ but $x(V \setminus X) = 0$ and so $x(Y) = 0$.
- Then $\mathrm{supp}(x) = X$, and say x is tight at X ($x(X) = f(X)$).

assert.

supp vs. sat equality

- For $x \in P_f$, with x extremal, is $\text{supp}(x) = \text{sat}(x)$?
- Consider an example case where disjoint $X, Y \subseteq E$, we have $f(X) = f(Y) = f(X \cup Y)$ (meaning "perfect dependence" or full redundancy, so gains are not strictly positive), $f(Y|X) = 0$.
- Suppose $x \in P_f$ has $x(X) > 0$ but $x(V \setminus X) = 0$ and so $x(Y) = 0$.
- Then $\text{supp}(x) = X$, and say x is tight at X ($x(X) = f(X)$).
- $\text{sat}(x) = \cup\{A : x(A) = f(A)\}$ and since $x(X \cup Y) = x(X) = f(X) = f(X \cup Y)$, here, $\text{sat}(x) \supseteq X \cup Y$.

$\therefore \text{sat}(x) > \text{supp}(x)$

supp vs. sat equality

- For $x \in P_f$, with x extremal, is $\text{supp}(x) = \text{sat}(x)$?
- Consider an example case where disjoint $X, Y \subseteq E$, we have $f(X) = f(Y) = f(X \cup Y)$ (meaning "perfect dependence" or full redundancy, so gains are not strictly positive), $f(Y|X) = 0$.
- Suppose $x \in P_f$ has $x(X) > 0$ but $x(V \setminus X) = 0$ and so $x(Y) = 0$.
- Then $\text{supp}(x) = X$, and say x is tight at X ($x(X) = f(X)$).
- $\text{sat}(x) = \cup\{A : x(A) = f(A)\}$ and since $x(X \cup Y) = x(X) = f(X) = f(X \cup Y)$, here, $\text{sat}(x) \supseteq X \cup Y$.
- In general, for extremal x, $\text{sat}(x) \supseteq \text{supp}(x)$ (see later).

supp vs. sat equality

- For $x \in P_f$, with x extremal, is $\text{supp}(x) = \text{sat}(x)$?
- Consider an example case where disjoint $X, Y \subseteq E$, we have $f(X) = f(Y) = f(X \cup Y)$ (meaning "perfect dependence" or full redundancy, so gains are not strictly positive), $f(Y|X) = 0$.
- Suppose $x \in P_f$ has $x(X) > 0$ but $x(V \setminus X) = 0$ and so $x(Y) = 0$.
- Then $\text{supp}(x) = X$, and say x is tight at X ($x(X) = f(X)$).
- $\text{sat}(x) = \cup\{A : x(A) = f(A)\}$ and since $x(X \cup Y) = x(X) = f(X) = f(X \cup Y)$, here, $\text{sat}(x) \supseteq X \cup Y$.
- In general, for extremal x, $\text{sat}(x) \supseteq \text{supp}(x)$ (see later).
- Also, recall $\text{sat}(x)$ is like span/closure but $\text{supp}(x)$ is more like indication. So this is similar to $\text{span}(A) \supseteq A$.

supp vs. sat equality

- For $x \in P_f$, with x extremal, is $\text{supp}(x) = \text{sat}(x)$?
- Consider an example case where disjoint $X, Y \subseteq E$, we have $f(X) = f(Y) = f(X \cup Y)$ (meaning "perfect dependence" or full redundancy, so gains are not strictly positive), $f(Y|X) = 0$.
- Suppose $x \in P_f$ has $x(X) > 0$ but $x(V \setminus X) = 0$ and so $x(Y) = 0$.
- Then $\text{supp}(x) = X$, and say x is tight at X ($x(X) = f(X)$).
- $\text{sat}(x) = \cup\{A : x(A) = f(A)\}$ and since $x(X \cup Y) = x(X) = f(X) = f(X \cup Y)$, here, $\text{sat}(x) \supseteq X \cup Y$.
- In general, for extremal x, $\text{sat}(x) \supseteq \text{supp}(x)$ (see later).
- Also, recall $\text{sat}(x)$ is like span/closure but $\text{supp}(x)$ is more like indication. So this is similar to $\text{span}(A) \supseteq A$.
- For modular functions, they are always equal at extreme points (e.g., think of "hyperrectangular" polymatroids).

Summary of supp, sat, and dep

- For $x \in P_f$, $\text{supp}(x) = \{e : x(e) \neq 0\} \subseteq \text{sat}(x)$
- For $x \in P_f$, $\text{sat}(x)$ (span, closure) is the maximal saturated (x-tight) set w.r.t. x. I.e., $\text{sat}(x) = \{e : e \in E, \forall \alpha > 0, x + \alpha \mathbf{1}_e \notin P_f\}$. That is,

$$\text{cl}(x) \stackrel{\text{def}}{=} \text{sat}(x) \triangleq \bigcup \{A : A \in \mathcal{D}(x)\} \tag{15.29}$$

$$= \bigcup \{A : A \subseteq E, x(A) = f(A)\} \tag{15.30}$$

$$= \{e : e \in E, \forall \alpha > 0, x + \alpha \mathbf{1}_e \notin P_f\} \tag{15.31}$$

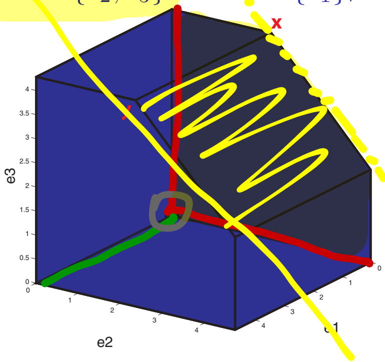
- For $e \in \text{sat}(x)$, we have $\text{dep}(x, e) \subseteq \text{sat}(x)$ (fundamental circuit) is the minimal (common) saturated (x-tight) set w.r.t. x containing e. I.e.,

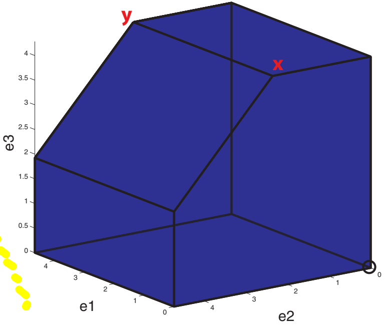
$$\text{dep}(x, e) = \begin{cases} \bigcap \{A : e \in A \subseteq E, x(A) = f(A)\} & \text{if } e \in \text{sat}(x) \\ \emptyset & \text{else} \end{cases}$$

$$= \{e' : \exists \alpha > 0, \text{ s.t. } x + \alpha(\mathbf{1}_e - \mathbf{1}_{e'}) \in P_f\} \tag{15.32}$$

supp, sat, dep, example with perfect independence

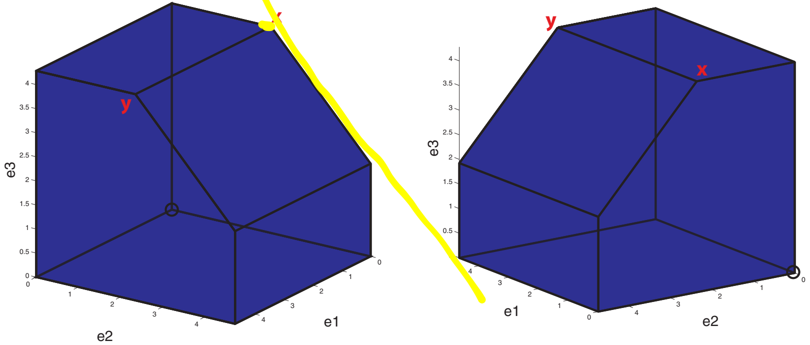
- Example polymatroid where there is perfect independence between $A = \{e_2, e_3\}$ and $B = \{e_1\}$, i.e., $e_1 \perp\!\!\!\perp \{e_2, e_3\}$.





supp, sat, dep, example with perfect independence

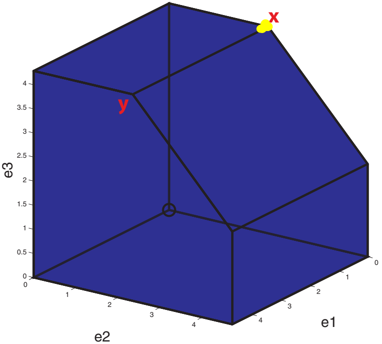
- Example polymatroid where there is perfect independence between $A = \{e_2, e_3\}$ and $B = \{e_1\}$, i.e., $e_1 \perp\!\!\!\perp \{e_2, e_3\}$.

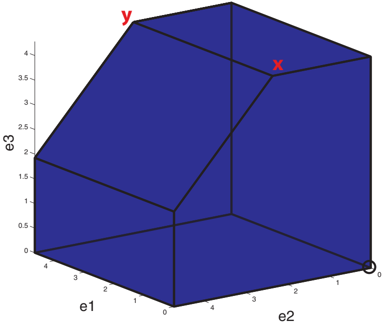


- Point x is extreme and $x(\{e_2, e_3\}) = f(e_2, e_3)$ (why?).

supp, sat, dep, example with perfect independence

- Example polymatroid where there is perfect independence between $A = \{e_2, e_3\}$ and $B = \{e_1\}$, i.e., $e_1 \perp\!\!\!\perp \{e_2, e_3\}$.

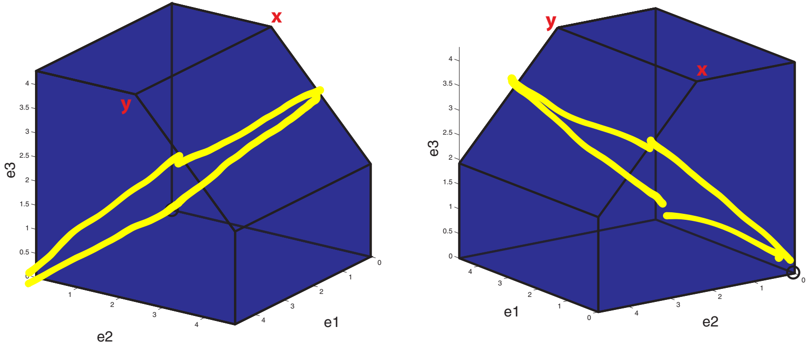




- Point x is extreme and $x(\{e_2, e_3\}) = f(e_2, e_3)$ (why?).
- But $x(\{e_1, e_2, e_3\}) = x(\{e_2, e_3\}) < f(e_1, e_2, e_3) = f(e_1) + f(e_2, e_3)$. Thus, $\text{supp}(x) = \text{sat}(x) = \{e_2, e_3\}$.

supp, sat, dep, example with perfect independence

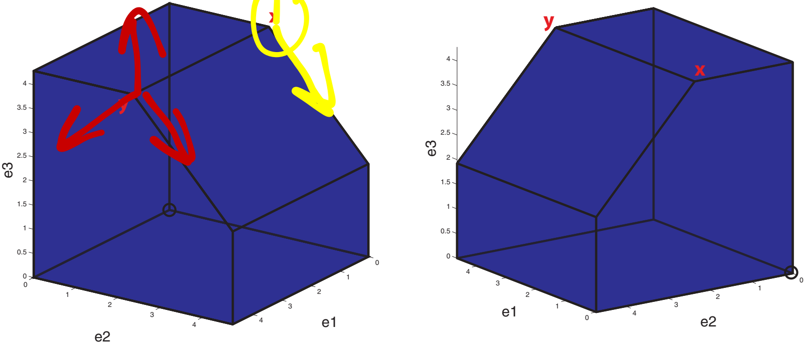
- Example polymatroid where there is perfect independence between $A = \{e_2, e_3\}$ and $B = \{e_1\}$, i.e., $e_1 \perp\!\!\!\perp \{e_2, e_3\}$.



- Note that considering a submodular function on clustered ground set $E = \{e_1, e_{23}\}$ where $f'(e_1) = f(e_1)$, $f'(e_{23}) = f(e_2, e_3)$ leads to a rectangle (no dependence between $\{e1\}$ and $\{e2, e3\}$).

supp, sat, dep, example with perfect independence

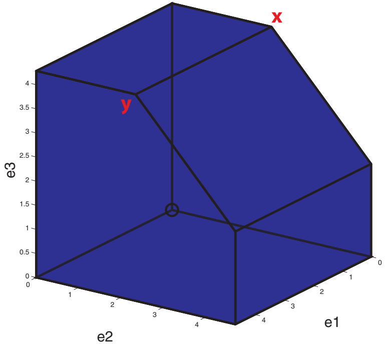
- Example polymatroid where there is perfect independence between $A = \{e_2, e_3\}$ and $B = \{e_1\}$, i.e., $e_1 \perp\!\!\!\perp \{e_2, e_3\}$.

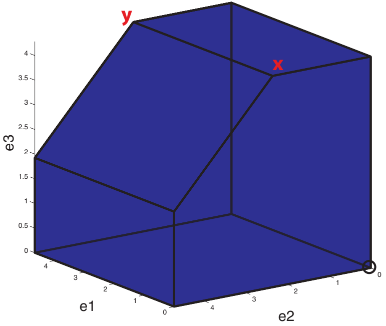


- We also have $\text{sat}(x) = \{e_3, e_2\}$. So $\text{dep}(x, e_1)$ is not defined, $\text{dep}(x, e_2) = \{e_3\}$, and $\text{dep}(x, e_3) = \emptyset$.
- $\text{sat}(y) = \{e_1, e_2, e_3\}$. So $\text{dep}(y, e_1) = \emptyset$, $\text{dep}(y, e_2) = e_3$, and $\text{dep}(y, e_3) = \emptyset$.

supp, sat, dep, example with perfect independence

- Example polymatroid where there is perfect independence between $A = \{e_2, e_3\}$ and $B = \{e_1\}$, i.e., $e_1 \perp\!\!\!\perp \{e_2, e_3\}$.

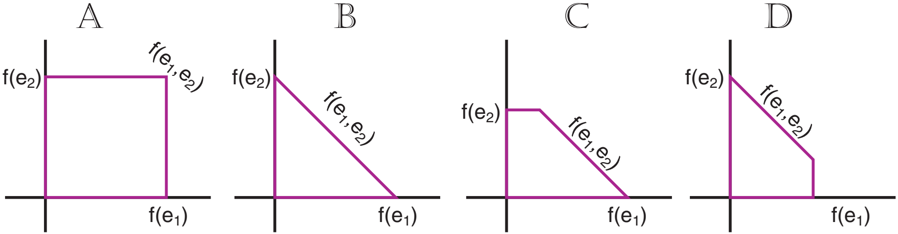




- We also have $\mathrm{sat}(x) = \{e_3, e_2\}$. So $\mathrm{dep}(x, e_1)$ is not defined, $\mathrm{dep}(x, e_2) = \{e_3\}$, and $\mathrm{dep}(x, e_3) = \emptyset$.
- $\mathrm{sat}(y) = \{e_1, e_2, e_3\}$. So $\mathrm{dep}(y, e_1) = \emptyset$, $\mathrm{dep}(y, e_2) = e_3$, and $\mathrm{dep}(y, e_3) = \emptyset$.

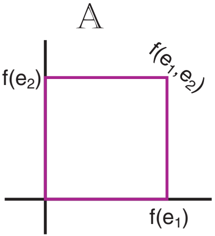
supp, sat, and polymatroid dependence in 2D

- Ex: various amounts of "dependence" between e_1 and e_2.



supp, sat, and polymatroid dependence in 2D

- Ex: various amounts of "dependence" between e_1 and e_2.

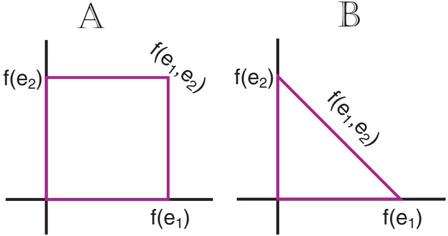


- Case A: perfect independence/irredunancy.

$f(e_1, e_2) = f(e_1) + f(e_2)$

supp, sat, and polymatroid dependence in 2D

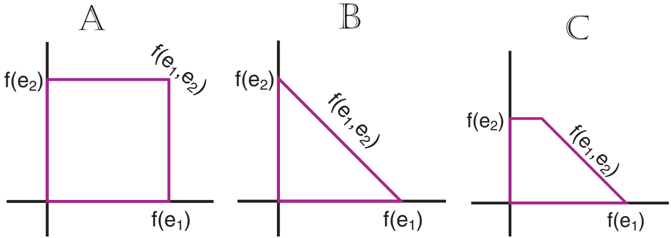
- Ex: various amounts of "dependence" between e_1 and e_2.



- Case A: perfect independence/irredunancy.
- Case B: perfect dependence/redundancy. Since slope is -45°, we must have $f(e_1) = f(e_2) = f(e_1, e_2)$. Entropy case: deterministic bijection between random variables e_1 and e_2.

supp, sat, and polymatroid dependence in 2D

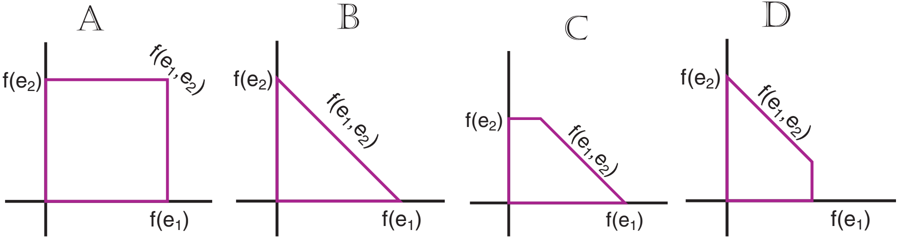
- Ex: various amounts of "dependence" between e_1 and e_2.



- Case A: perfect independence/irredunancy.
- Case B: perfect dependence/redundancy. Since slope is -45°, we must have $f(e_1) = f(e_2) = f(e_1, e_2)$. Entropy case: deterministic bijection between random variables e_1 and e_2.
- Case C: $f(e_2) < f(e_1) = f(e_1, e_2)$. Entropy case: random variable e_2 a deterministic function of e_1 which has higher entropy.

supp, sat, and polymatroid dependence in 2D

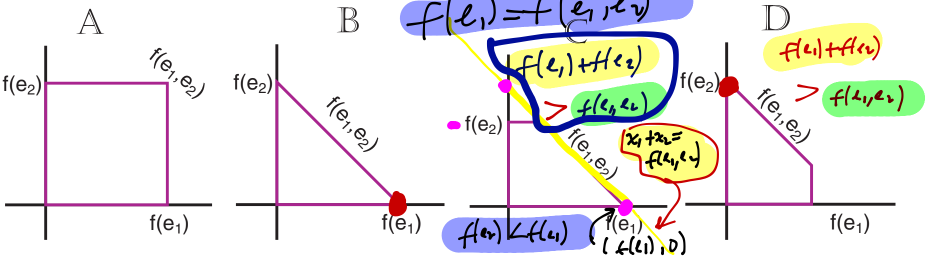
- Ex: various amounts of "dependence" between e_1 and e_2.



- Case A: perfect independence/irredunancy.
- Case B: perfect dependence/redundancy. Since slope is -45°, we must have $f(e_1) = f(e_2) = f(e_1, e_2)$. Entropy case: deterministic bijection between random variables e_1 and e_2.
- Case C: $f(e_2) < f(e_1) = f(e_1, e_2)$. Entropy case: random variable e_2 a deterministic function of e_1 which has higher entropy.
- Case D: $f(e_1) < f(e_2) = f(e_1, e_2)$. Entropy case: random variable e_1 a deterministic function of e_2 which has higher entropy.

supp, sat, and polymatroid dependence in 2D

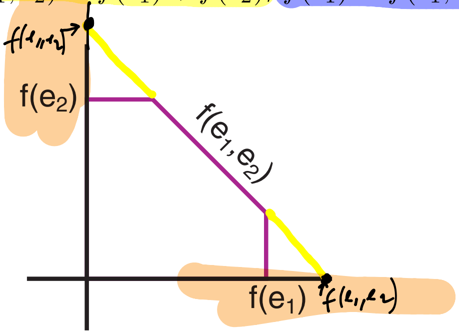
- Ex: various amounts of "dependence" between e_1 and e_2.



- In each case, we see points x where $\text{supp}(x) \subseteq \text{sat}(x)$.
- Example: Case B or C, let $x = (f(e_1), 0)$ so $\text{supp}(x) = \{e_1\}$ but since $x(\{e_1, e_2\}) = x(\{e_1\}) = f(e_1) = f(e_1, e_2)$ we have $\text{sat}(x) = \{e_1, e_2\}$.
- Similar for case D with $x = (0, f(e_2))$.

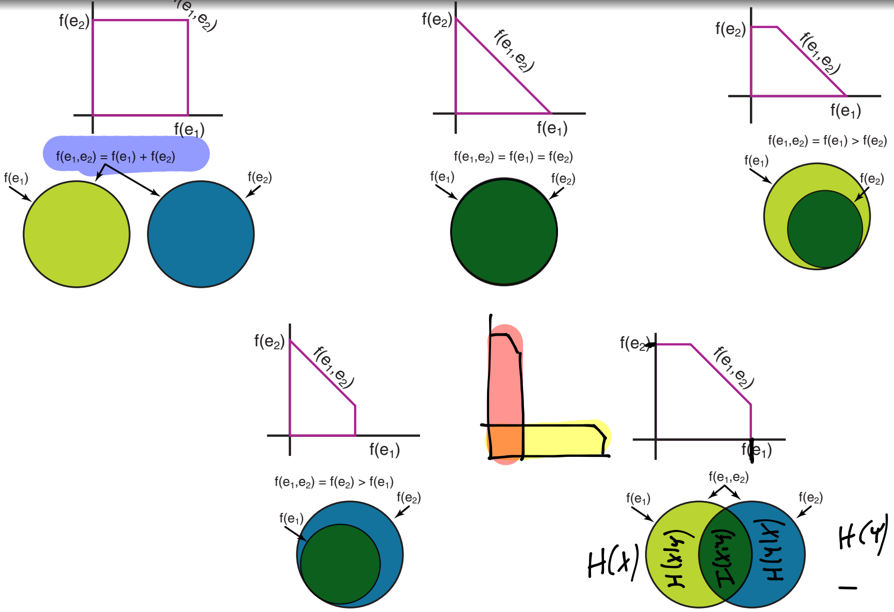
supp, sat, and dependence in 2D

- General case, $f(e_1, e_2) < f(e_1) + f(e_2)$, $f(e_1) < f(e_1, e_2)$, and $f(e_2) < f(e_1, e_2)$.



- Entropy case: We have a random variable Z and two separate deterministic functions $e_1 = h_1(Z)$ and $e_2 = h_2(Z)$ such that the entropy $H(e_1, e_2) = H(Z)$, but each deterministic function gives a different "view" of Z, each contains more than half the information, and the two are redundant w.r.t. each other $(H(e_1) + H(e_2) > H(Z))$.

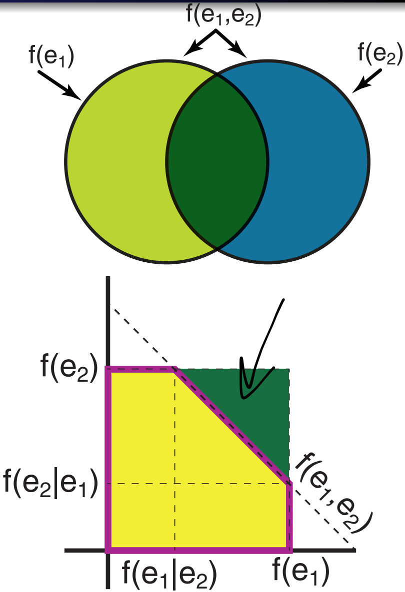
2D polymatroids and Information Venn Diagrams



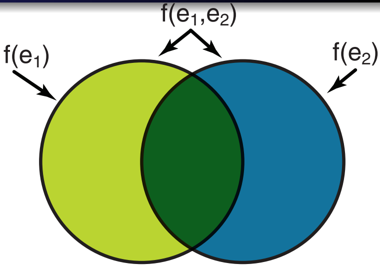
2D polymatroids and Information Venn Diagrams

- Consider symmeterized combinatorial mutual information function:

$$I_f(e_1, e_2) = f(e_1) + f(e_2) - f(e_1, e_2)$$



2D polymatroids and Information Venn Diagrams

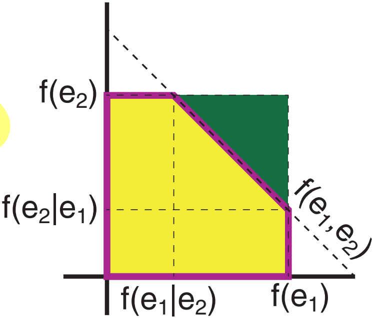


- Consider symmeterized combinatorial mutual information function:

$$I_f(e_1, e_2) = f(e_1) + f(e_2) - f(e_1, e_2)$$

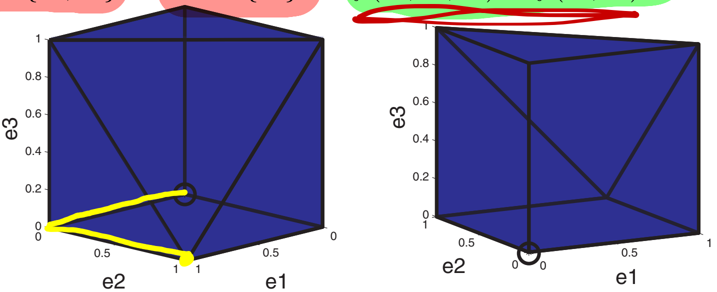
- Consider area of green triangle:

$$\frac{1}{2}\Big(f(e_2) - f(e_2|e_1)\Big)\Big(f(e_1) - f(e_1|e_2)\Big)$$
$$= \frac{1}{2}(f(e_1) + f(e_2) - f(e_1, e_2))^2$$
$$= \frac{1}{2}\Big(I_f(e_1, e_2)\Big)^2$$



supp, sat, and perfect dependence in 3D

- Ex: polymatroid with perfect independence between e_2 and e_3, so $f(e_2, e_3) = f(e_2) + f(e_3)$, but perfect dependence between $A = \{e_2, e_3\}$ and $B = \{e_1\}$, so $f(e_1, e_2, e_3) = f(e_2, e_3)$

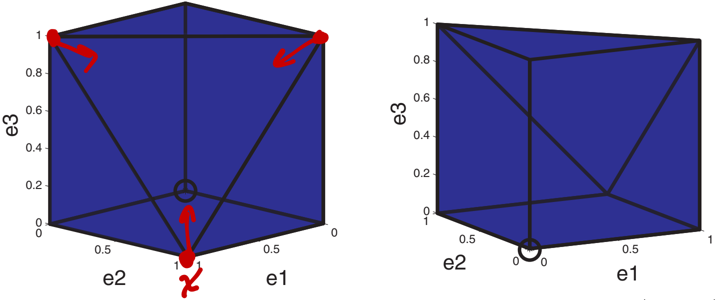


$f(e_3 | e_1, e_2) = 0$

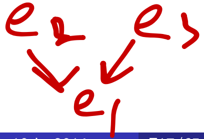
$f(e_1, e_2) = f(e_1, e_2, e_3)$

supp, sat, and perfect dependence in 3D

- Ex: polymatroid with perfect independence between e_2 and e_3, so $f(e_2, e_3) = f(e_2) + f(e_3)$, but perfect dependence between $A = \{e_2, e_3\}$ and $B = \{e_1\}$, so $f(e_1, e_2, e_3) = f(e_2, e_3)$

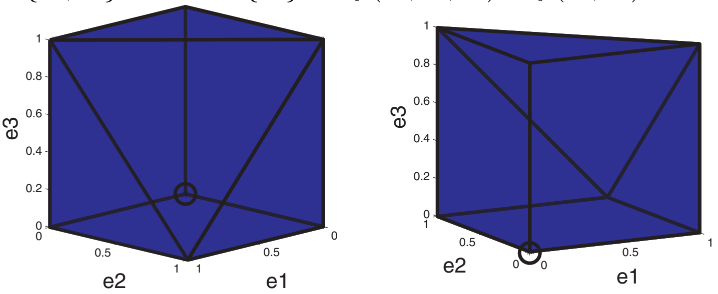


- Entropy case: xor V-structure Bayesian network $e_1 = h(e_2, e_3)$ where h is the xor function ($e_2 \to e_1 \leftarrow e_3$), and e_2, e_3 are both independent binary with unity entropy.



supp, sat, and perfect dependence in 3D

- Ex: polymatroid with perfect independence between e_2 and e_3, so $f(e_2, e_3) = f(e_2) + f(e_3)$, but perfect dependence between $A = \{e_2, e_3\}$ and $B = \{e_1\}$, so $f(e_1, e_2, e_3) = f(e_2, e_3)$



- Entropy case: xor V-structure Bayesian network $e_1 = h(e_2, e_3)$ where h is the xor function $(e_2 \rightarrow e_1 \leftarrow e_3)$, and e_2, e_3 are both independent binary with unity entropy.
- Q: Why does the polytope have a symmetry? Notice independence (square) for any pair.

\begin{tikzcd}
e_2 \arrow[dr] & & e_3 \arrow[dl] \\
& e_1 &
\end{tikzcd}

$$e_2 \nparallel e_3$$

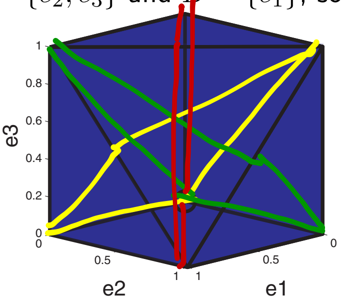
$$e_1 = e_2 \text{ xor } e_3$$

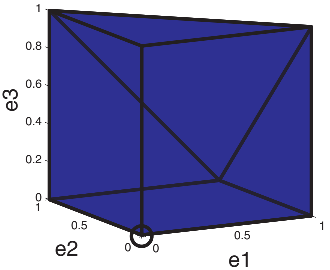
$$\equiv$$

$$
\begin{tikzcd}
e_1 \arrow[dr] & & e_2 \arrow[dl] \\
& e_3 &
\end{tikzcd}
\quad \equiv \quad
\begin{tikzcd}
e_1 \arrow[dr] & & e_3 \arrow[dl] \\
& e_2 &
\end{tikzcd}
$$

supp, sat, and perfect dependence in 3D

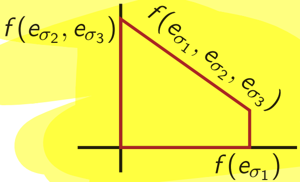
- Ex: polymatroid with perfect independence between e_2 and e_3, so $f(e_2, e_3) = f(e_2) + f(e_3)$, but perfect dependence between $A = \{e_2, e_3\}$ and $B = \{e_1\}$, so $f(e_1, e_2, e_3) = f(e_2, e_3)$





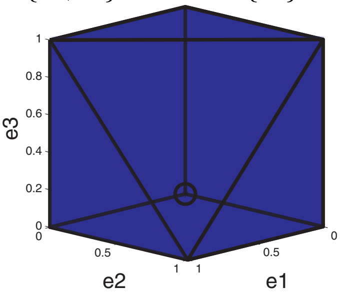
- For any permutation σ of $\{1, 2, 3\}$, considering $\{e_{\sigma_1}, e_{\sigma_2}\}$ vs. $\{e_{\sigma_3}\}$:

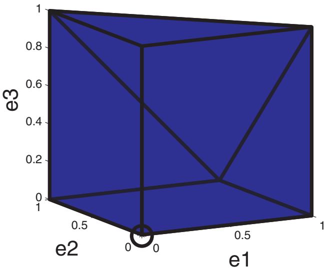
e_{σ_3} is a deterministic function of $\{e_{\sigma_1}, e_{\sigma_2}\}$



supp, sat, and perfect dependence in 3D

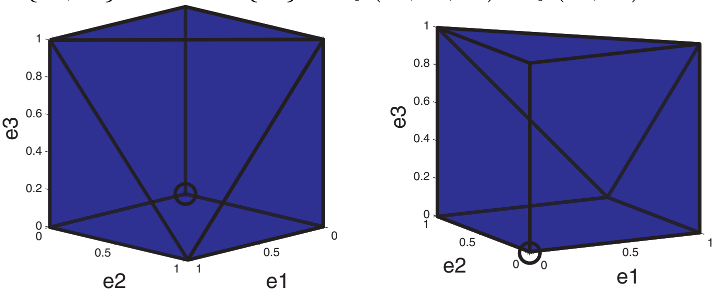
- Ex: polymatroid with perfect independence between e_2 and e_3, so $f(e_2, e_3) = f(e_2) + f(e_3)$, but perfect dependence between $A = \{e_2, e_3\}$ and $B = \{e_1\}$, so $f(e_1, e_2, e_3) = f(e_2, e_3)$





supp, sat, and perfect dependence in 3D

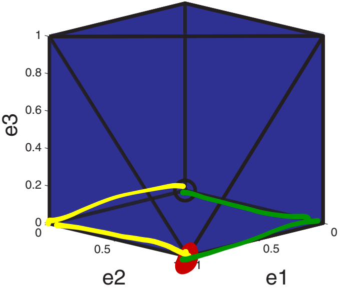
- Ex: polymatroid with perfect independence between e_2 and e_3, so $f(e_2, e_3) = f(e_2) + f(e_3)$, but perfect dependence between $A = \{e_2, e_3\}$ and $B = \{e_1\}$, so $f(e_1, e_2, e_3) = f(e_2, e_3)$

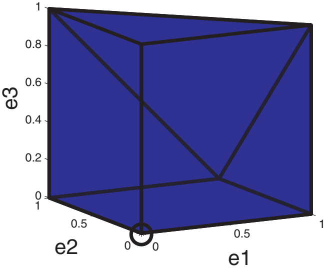


- Note also, that for some of the extreme points, multiple orders generate them.

supp, sat, and perfect dependence in 3D

- Ex: polymatroid with perfect independence between e_2 and e_3, so $f(e_2, e_3) = f(e_2) + f(e_3)$, but perfect dependence between $A = \{e_2, e_3\}$ and $B = \{e_1\}$, so $f(e_1, e_2, e_3) = f(e_2, e_3)$





- Note also, that for some of the extreme points, multiple orders generate them.
- Consider extreme point $x = (x_1, x_2, x_3) = (1, 1, 0)$. Then we get this either with orders (e_1, e_2, e_3), or (e_2, e_1, e_3). This is true since $f(e_{\sigma_e} | \{e_{\sigma_1}, e_{\sigma_2}\}) = 0$ for all permutations σ of $\{1, 2, 3\}$.

perfect dependence in 3D, entropy, and Bayesian networks

- The example in the previous slides can be realized with entropy of random variables and a Bayesian network.

perfect dependence in 3D, entropy, and Bayesian networks

- The example in the previous slides can be realized with entropy of random variables and a Bayesian network.
- Consider three binary random variables $X_1, X_2, X_3 \in \{0, 1\}$ that factor w.r.t., the V-structure $X_1 \rightarrow X_3 \leftarrow X_2$, where $X_3 = X_1 \oplus X_2$, where $\oplus$ is the X-OR operator, and where $X_1 \perp\!\!\!\perp X_2$.

perfect dependence in 3D, entropy, and Bayesian networks

- The example in the previous slides can be realized with entropy of random variables and a Bayesian network.
- Consider three binary random variables $X_1, X_2, X_3 \in \{0, 1\}$ that factor w.r.t., the V-structure $X_1 \to X_3 \leftarrow X_2$, where $X_3 = X_1 \oplus X_2$, where $\oplus$ is the X-OR operator, and where $X_1 \perp\!\!\!\perp X_2$.
- Consequently, $X_i \perp\!\!\!\perp X_j$ for any $i \neq j$.

perfect dependence in 3D, entropy, and Bayesian networks

- The example in the previous slides can be realized with entropy of random variables and a Bayesian network.
- Consider three binary random variables $X_1, X_2, X_3 \in \{0, 1\}$ that factor w.r.t., the V-structure $X_1 \rightarrow X_3 \leftarrow X_2$, where $X_3 = X_1 \oplus X_2$, where $\oplus$ is the X-OR operator, and where $X_1 \perp\!\!\!\perp X_2$.
- Consequently, $X_i \perp\!\!\!\perp X_j$ for any $i \neq j$.
- Moreover, for any permutation σ of $\{1, 2, 3\}$, we have the relationship $X_{\sigma_1} = X_{\sigma_2} \oplus X_{\sigma_3}$.

perfect dependence in 3D, entropy, and Bayesian networks

- The example in the previous slides can be realized with entropy of random variables and a Bayesian network.
- Consider three binary random variables $X_1, X_2, X_3 \in \{0, 1\}$ that factor w.r.t., the V-structure $X_1 \rightarrow X_3 \leftarrow X_2$, where $X_3 = X_1 \oplus X_2$, where $\oplus$ is the X-OR operator, and where $X_1 \perp\!\!\!\perp X_2$.
- Consequently, $X_i \perp\!\!\!\perp X_j$ for any $i \neq j$.
- Moreover, for any permutation σ of $\{1, 2, 3\}$, we have the relationship $X_{\sigma_1} = X_{\sigma_2} \oplus X_{\sigma_3}$.
- The entropy function $f(A) = H(X_A)$ is a submodular function that will have the symmetric 3D polytope of the previous example.

supp, sat, extremal x, perfect dependence

- In general, for extremal x, $\text{sat}(x) \supseteq \text{supp}(x)$.

supp, sat, extremal x, perfect dependence

- In general, for extremal x, $\text{sat}(x) \supseteq \text{supp}(x)$.
- Now, for any $e \in E \setminus \text{supp}(x)$, we clearly have $x(\text{supp}(x) + e) = x(\text{supp}(x))$ since $x(e) = 0$.

supp, sat, extremal x, perfect dependence

- In general, for extremal x, $\text{sat}(x) \supseteq \text{supp}(x)$.
- Now, for any $e \in E \setminus \text{supp}(x)$, we clearly have $x(\text{supp}(x) + e) = x(\text{supp}(x))$ since $x(e) = 0$.
- On the other hand, for $e_i \in \text{sat}(x) \setminus \text{supp}(x)$, we have perfect dependence, i.e., $f(\text{supp}(x) + e_i) = f(\text{supp}(x))$. Proof:

$$f(e_i | \text{supp}(x)) \leq 0$$

supp, sat, extremal x, perfect dependence

- In general, for extremal x, $\text{sat}(x) \supseteq \text{supp}(x)$.
- Now, for any $e \in E \setminus \text{supp}(x)$, we clearly have $x(\text{supp}(x) + e) = x(\text{supp}(x))$ since $x(e) = 0$.
- On the other hand, for $e_i \in \text{sat}(x) \setminus \text{supp}(x)$, we have perfect dependence, i.e., $f(\text{supp}(x) + e_i) = f(\text{supp}(x))$. Proof:
    - $\text{sat}(x)$ is tight, as is $\text{supp}(x)$, and hence $f(\text{sat}(x)) = x(\text{sat}(x)) = x(\text{supp}(x)) = f(\text{supp}(x))$.

supp, sat, extremal x, perfect dependence

- In general, for extremal x, $\mathrm{sat}(x) \supseteq \mathrm{supp}(x)$.
- Now, for any $e \in E \setminus \mathrm{supp}(x)$, we clearly have $x(\mathrm{supp}(x) + e) = x(\mathrm{supp}(x))$ since $x(e) = 0$.
- On the other hand, for $e_i \in \mathrm{sat}(x) \setminus \mathrm{supp}(x)$, we have perfect dependence, i.e., $f(\mathrm{supp}(x) + e_i) = f(\mathrm{supp}(x))$. Proof:
  - $\mathrm{sat}(x)$ is tight, as is $\mathrm{supp}(x)$, and hence $f(\mathrm{sat}(x)) = x(\mathrm{sat}(x)) = x(\mathrm{supp}(x)) = f(\mathrm{supp}(x))$.
  - Therefore, $f(\mathrm{sat}(x) | \mathrm{supp}(x)) = 0$.

supp, sat, extremal x, perfect dependence

- In general, for extremal x, $\text{sat}(x) \supseteq \text{supp}(x)$.
- Now, for any $e \in E \setminus \text{supp}(x)$, we clearly have $x(\text{supp}(x) + e) = x(\text{supp}(x))$ since $x(e) = 0$.
- On the other hand, for $e_i \in \text{sat}(x) \setminus \text{supp}(x)$, we have perfect dependence, i.e., $f(\text{supp}(x) + e_i) = f(\text{supp}(x))$. Proof:
  - $\text{sat}(x)$ is tight, as is $\text{supp}(x)$, and hence $f(\text{sat}(x)) = x(\text{sat}(x)) = x(\text{supp}(x)) = f(\text{supp}(x))$.
  - Therefore, $f(\text{sat}(x) | \text{supp}(x)) = 0$.
  - But by the above, and monotonicity, we have for $e \in \text{sat}(x) \setminus \text{supp}(x)$, that $0 = f(\text{sat}(x) | \text{supp}(x)) \geq f(e | \text{supp}(x)) \geq 0$.

supp, sat, extremal x, perfect dependence

- In general, for extremal x, $\text{sat}(x) \supseteq \text{supp}(x)$.
- Now, for any $e \in E \setminus \text{supp}(x)$, we clearly have $x(\text{supp}(x) + e) = x(\text{supp}(x))$ since $x(e) = 0$.
- On the other hand, for $e_i \in \text{sat}(x) \setminus \text{supp}(x)$, we have perfect dependence, i.e., $f(\text{supp}(x) + e_i) = f(\text{supp}(x))$. Proof:
    - $\text{sat}(x)$ is tight, as is $\text{supp}(x)$, and hence $f(\text{sat}(x)) = x(\text{sat}(x)) = x(\text{supp}(x)) = f(\text{supp}(x))$.
    - Therefore, $f(\text{sat}(x) | \text{supp}(x)) = 0$.
    - But by the above, and monotonicity, we have for $e \in \text{sat}(x) \setminus \text{supp}(x)$, that $0 = f(\text{sat}(x) | \text{supp}(x)) \geq f(e | \text{supp}(x)) \geq 0$.
    - Hence $f(e | \text{supp}(x)) = 0$,

supp, sat, extremal x, perfect dependence

- In general, for extremal x, $\text{sat}(x) \supseteq \text{supp}(x)$.
- Now, for any $e \in E \setminus \text{supp}(x)$, we clearly have $x(\text{supp}(x) + e) = x(\text{supp}(x))$ since $x(e) = 0$.
- On the other hand, for $e_i \in \text{sat}(x) \setminus \text{supp}(x)$, we have perfect dependence, i.e., $f(\text{supp}(x) + e_i) = f(\text{supp}(x))$. Proof:
    - $\text{sat}(x)$ is tight, as is $\text{supp}(x)$, and hence $f(\text{sat}(x)) = x(\text{sat}(x)) = x(\text{supp}(x)) = f(\text{supp}(x))$.
    - Therefore, $f(\text{sat}(x)|\,\text{supp}(x)) = 0$.
    - But by the above, and monotonicity, we have for $e \in \text{sat}(x) \setminus \text{supp}(x)$, that $0 = f(\text{sat}(x)|\,\text{supp}(x)) \geq f(e|\,\text{supp}(x)) \geq 0$.
    - Hence $f(e|\,\text{supp}(x)) = 0$, and moreover $f(e + \text{supp}(x)) = x(e + \text{supp}(x)) = x(\text{supp}(x)) = f(\text{supp}(x))$.

supp, sat, extremal x, perfect dependence

- In general, for extremal x, $\mathrm{sat}(x) \supseteq \mathrm{supp}(x)$.
- Now, for any $e \in E \setminus \mathrm{supp}(x)$, we clearly have $x(\mathrm{supp}(x) + e) = x(\mathrm{supp}(x))$ since $x(e) = 0$.
- On the other hand, for $e_i \in \mathrm{sat}(x) \setminus \mathrm{supp}(x)$, we have perfect dependence, i.e., $f(\mathrm{supp}(x) + e_i) = f(\mathrm{supp}(x))$. Proof:
  - $\mathrm{sat}(x)$ is tight, as is $\mathrm{supp}(x)$, and hence $f(\mathrm{sat}(x)) = x(\mathrm{sat}(x)) = x(\mathrm{supp}(x)) = f(\mathrm{supp}(x))$.
  - Therefore, $f(\mathrm{sat}(x)|\,\mathrm{supp}(x)) = 0$.
  - But by the above, and monotonicity, we have for $e \in \mathrm{sat}(x) \setminus \mathrm{supp}(x)$, that $0 = f(\mathrm{sat}(x)|\,\mathrm{supp}(x)) \geq f(e|\,\mathrm{supp}(x)) \geq 0$.
  - Hence $f(e|\,\mathrm{supp}(x)) = 0$, and moreover $f(e + \mathrm{supp}(x)) = x(e + \mathrm{supp}(x)) = x(\mathrm{supp}(x)) = f(\mathrm{supp}(x))$.
- Thus, for any extremal x, with $\mathrm{sat}(x) \supset \mathrm{supp}(x)$, we see that for $e \in \mathrm{sat}(x) \setminus \mathrm{supp}(x)$, we have $\mathrm{supp}(x) + e$ is also tight.

supp, sat, extremal x, perfect dependence

- In general, for extremal x, $\text{sat}(x) \supseteq \text{supp}(x)$.
- Now, for any $e \in E \setminus \text{supp}(x)$, we clearly have $x(\text{supp}(x) + e) = x(\text{supp}(x))$ since $x(e) = 0$.
- On the other hand, for $e_i \in \text{sat}(x) \setminus \text{supp}(x)$, we have perfect dependence, i.e., $f(\text{supp}(x) + e_i) = f(\text{supp}(x))$. Proof:
    - $\text{sat}(x)$ is tight, as is $\text{supp}(x)$, and hence $f(\text{sat}(x)) = x(\text{sat}(x)) = x(\text{supp}(x)) = f(\text{supp}(x))$.
    - Therefore, $f(\text{sat}(x)|\,\text{supp}(x)) = 0$.
    - But by the above, and monotonicity, we have for $e \in \text{sat}(x) \setminus \text{supp}(x)$, that $0 = f(\text{sat}(x)|\,\text{supp}(x)) \geq f(e|\,\text{supp}(x)) \geq 0$.
    - Hence $f(e|\,\text{supp}(x)) = 0$, and moreover $f(e + \text{supp}(x)) = x(e + \text{supp}(x)) = x(\text{supp}(x)) = f(\text{supp}(x))$.
- Thus, for any extremal x, with $\text{sat}(x) \supset \text{supp}(x)$, we see that for $e \in \text{sat}(x) \setminus \text{supp}(x)$, we have $\text{supp}(x) + e$ is also tight.
- Note also, for any $A \subseteq \text{sat}(x) \setminus \text{supp}(x)$, we have $f(A|\,\text{supp}(x)) = 0$.

supp, sat, perfect dependence

- Note that all of these results hold when f is monotone non-decreasing submodular (e.g., for a polymatroid function).

supp, sat, perfect dependence

- Note that all of these results hold when f is monotone non-decreasing submodular (e.g., for a polymatroid function).
- For modular functions, and extremal x, $sat(x) = \text{supp}(x)$.

supp, sat, perfect dependence

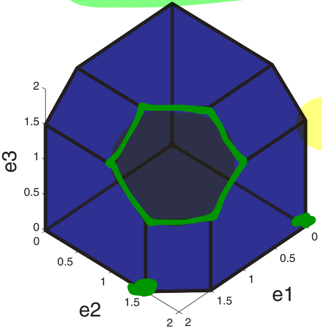
- Note that all of these results hold when f is monotone non-decreasing submodular (e.g., for a polymatroid function).
- For modular functions, and extremal x, $sat(x) = \text{supp}(x)$.
- For general $x \in P_f$ (not nec. extremal), $sat(x)$ and $\text{supp}(x)$ might have an arbitrary relationship (but we want to strengthen this relationship further, and we will do so below).

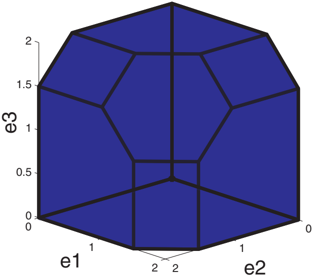
supp, sat, perfect dependence

- Note that all of these results hold when f is monotone non-decreasing submodular (e.g., for a polymatroid function).
- For modular functions, and extremal x, $sat(x) = \text{supp}(x)$.
- For general $x \in P_f$ (not nec. extremal), $sat(x)$ and $\text{supp}(x)$ might have an arbitrary relationship (but we want to strengthen this relationship further, and we will do so below).
- For the most part, we are interested in these quantities when x is extremal as we will see.

supp and sat, example under limited curvature

- Strict monotone f polymatroids, where $f(e|E \setminus e) > 0, \forall e$.
- Example: $f(A) = \sqrt{|A|}$, where all $m!$ vertices of B_f are unique.





- In such cases, taking any extremal point $x \in P_f$ based on prefix order $E = (e_1, \ldots)$, where $\text{supp}(x) \subset E$, we have that $\text{sat}(x) = \text{supp}(x)$ since the largest tight set corresponds to $x(E_i) = f(E_i)$ for some i, and while any $e \in E \setminus E_i$ is such that $x(E_i + e) = x(E_i)$, there is **no** such e with $f(E_i + e) = f(E_i)$.

Another revealing theorem

Theorem 15.5.1

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

Another revealing theorem

Theorem 15.5.1

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

- Thus, "independence" between disjoint A and B (leading to a rectangular projection of the polymatroid polytope) reduces the dimension of the base polytope, as expected.

Another revealing theorem

Theorem 15.5.1

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

- Thus, "independence" between disjoint A and B (leading to a rectangular projection of the polymatroid polytope) reduces the dimension of the base polytope, as expected.
- Thus, any point $x \in B_f$ is a convex combination of at most $|E| - k + 1$ vertices of B_f.

Another revealing theorem

Theorem 15.5.1

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

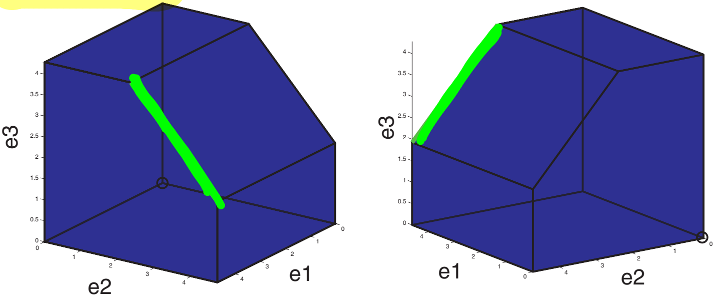
- Thus, "independence" between disjoint A and B (leading to a rectangular projection of the polymatroid polytope) reduces the dimension of the base polytope, as expected.
- Thus, any point $x \in B_f$ is a convex combination of at most $|E| - k + 1$ vertices of B_f.
- And if f does not have such independence, dimension of B_f is $|E| - 1$ and any point $x \in B_f$ is a convex combination of at most $|E|$ vertices of B_f.

Another revealing theorem

Theorem 15.5.1

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

- Example f with independence between $A = \{e_2, e_3\}$ and $B = \{e_1\}$, i.e., $e_1 \perp\!\!\!\perp \{e_2, e_3\}$, with B_f marked in green.



Base polytope existence

- Given polymatroid function f, the base polytope $B_f = \{x \in \mathbb{R}_+^E : x(A) \leq f(A) \; \forall A \subseteq E, \text{ and } x(E) = f(E)\}$ always exists.

Base polytope existence

- Given polymatroid function f, the base polytope $B_f = \left\{x \in \mathbb{R}_+^E : x(A) \leq f(A) \; \forall A \subseteq E, \text{ and } x(E) = f(E)\right\}$ always exists.
- Consider any order of E and generate a vector x by this order (i.e., $x(e_1) = f(\{e_1\})$, $x(e_2) = f(\{e_1, e_2\}) - f(\{e_1\})$, and so on).

Base polytope existence

- Given polymatroid function f, the base polytope $B_f = \{x \in \mathbb{R}_+^E : x(A) \leq f(A)\ \forall A \subseteq E, \text{ and } x(E) = f(E)\}$ always exists.
- Consider any order of E and generate a vector x by this order (i.e., $x(e_1) = f(\{e_1\})$, $x(e_2) = f(\{e_1, e_2\}) - f(\{e_1\})$, and so on).
- From past lectures, we now know that:

Base polytope existence

- Given polymatroid function f, the base polytope $B_f = \left\{x \in \mathbb{R}_+^E : x(A) \leq f(A)\ \forall A \subseteq E, \text{ and } x(E) = f(E)\right\}$ always exists.
- Consider any order of E and generate a vector x by this order (i.e., $x(e_1) = f(\{e_1\})$, $x(e_2) = f(\{e_1, e_2\}) - f(\{e_1\})$, and so on).
- From past lectures, we now know that:

  (1) $x \in P_f$

Base polytope existence

- Given polymatroid function f, the base polytope $B_f = \left\{ x \in \mathbb{R}_+^E : x(A) \le f(A)\ \forall A \subseteq E, \text{ and } x(E) = f(E) \right\}$ always exists.
- Consider any order of E and generate a vector x by this order (i.e., $x(e_1) = f(\{e_1\})$, $x(e_2) = f(\{e_1, e_2\}) - f(\{e_1\})$, and so on).
- From past lectures, we now know that:
  (1) $x \in P_f$
  (2) x is an extreme point in P_f

Base polytope existence

- Given polymatroid function f, the base polytope $B_f = \left\{x \in \mathbb{R}_+^E : x(A) \leq f(A)\ \forall A \subseteq E,\ \text{and } x(E) = f(E)\right\}$ always exists.
- Consider any order of E and generate a vector x by this order (i.e., $x(e_1) = f(\{e_1\})$, $x(e_2) = f(\{e_1, e_2\}) - f(\{e_1\})$, and so on).
- From past lectures, we now know that:
  1. $x \in P_f$
  2. x is an extreme point in P_f
  3. Since x is generated using an ordering of all of E, we have that $x(E) = f(E)$.

Base polytope existence

- Given polymatroid function f, the base polytope $B_f = \{x \in \mathbb{R}_+^E : x(A) \leq f(A)\ \forall A \subseteq E,\ \text{and } x(E) = f(E)\}$ always exists.
- Consider any order of E and generate a vector x by this order (i.e., $x(e_1) = f(\{e_1\})$, $x(e_2) = f(\{e_1, e_2\}) - f(\{e_1\})$, and so on).
- From past lectures, we now know that:
  1. $x \in P_f$
  2. x is an extreme point in P_f
  3. Since x is generated using an ordering of all of E, we have that $x(E) = f(E)$.
- Thus $x \in B_f$, and B_f is never empty.

Base polytope existence

- Given polymatroid function f, the base polytope $B_f = \left\{x \in \mathbb{R}_+^E : x(A) \leq f(A)\ \forall A \subseteq E, \text{ and } x(E) = f(E)\right\}$ always exists.
- Consider any order of E and generate a vector x by this order (i.e., $x(e_1) = f(\{e_1\})$, $x(e_2) = f(\{e_1, e_2\}) - f(\{e_1\})$, and so on).
- From past lectures, we now know that:
  (1) $x \in P_f$
  (2) x is an extreme point in P_f
  (3) Since x is generated using an ordering of all of E, we have that $x(E) = f(E)$.
- Thus $x \in B_f$, and B_f is never empty.
- Moreover, in this case, x is a vertex of B_f since it is extremal.

Base polytope property

- Now, for any $A \subseteq E$, we can generate a particular point in B_f

$$x \in B_f \subseteq P_f$$

$$x(A) = f(A)$$

$$x(B) \leq f(B) \quad \forall B.$$

Base polytope property

- Now, for any $A \subseteq E$, we can generate a particular point in B_f
- That is, choose the ordering of $E = (e_1, e_2, \ldots, e_n)$ where $n = |E|$, and where $E_i = (e_1, e_2, \ldots, e_i)$, so that we have $E_k = A$ with $k = |A|$.

Base polytope property

- Now, for any $A \subseteq E$, we can generate a particular point in B_f
- That is, choose the ordering of $E = (e_1, e_2, \ldots, e_n)$ where $n = |E|$, and where $E_i = (e_1, e_2, \ldots, e_i)$, so that we have $E_k = A$ with $k = |A|$.
- Note there are $k!(n-k)! < n!$ such orderings.

Base polytope property

- Now, for any $A \subseteq E$, we can generate a particular point in B_f
- That is, choose the ordering of $E = (e_1, e_2, \ldots, e_n)$ where $n = |E|$, and where $E_i = (e_1, e_2, \ldots, e_i)$, so that we have $E_k = A$ with $k = |A|$.
- Note there are $k!(n-k)! < n!$ such orderings.
- Generate x via greedy using this order, $\forall i, x(e_i) = f(e_i | E_{i-1})$.

Base polytope property

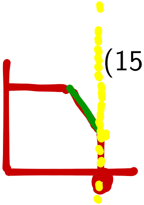
- Now, for any $A \subseteq E$, we can generate a particular point in B_f
- That is, choose the ordering of $E = (e_1, e_2, \dots, e_n)$ where $n = |E|$, and where $E_i = (e_1, e_2, \dots, e_i)$, so that we have $E_k = A$ with $k = |A|$.
- Note there are $k!(n-k)! < n!$ such orderings.
- Generate x via greedy using this order, $\forall i, x(e_i) = f(e_i | E_{i-1})$.
- Then, we have generated a point x (a vertex, no less) in B_f such that $x(A) = f(A)$.

Base polytope property

- Now, for any $A \subseteq E$, we can generate a particular point in B_f
- That is, choose the ordering of $E = (e_1, e_2, \dots, e_n)$ where $n = |E|$, and where $E_i = (e_1, e_2, \dots, e_i)$, so that we have $E_k = A$ with $k = |A|$.
- Note there are $k!(n-k)! < n!$ such orderings.
- Generate x via greedy using this order, $\forall i, x(e_i) = f(e_i | E_{i-1})$.
- Then, we have generated a point x (a vertex, no less) in B_f such that $x(A) = f(A)$.
- Thus, for any A, we have

$$B_f \cap \{x \in \mathbb{R}^E : x(A) = f(A)\} \neq \emptyset \tag{15.2}$$

$x(E)=f(E)$



Base polytope property

- Now, for any $A \subseteq E$, we can generate a particular point in B_f
- That is, choose the ordering of $E = (e_1, e_2, \ldots, e_n)$ where $n = |E|$, and where $E_i = (e_1, e_2, \ldots, e_i)$, so that we have $E_k = A$ with $k = |A|$.
- Note there are $k!(n-k)! < n!$ such orderings.
- Generate x via greedy using this order, $\forall i, x(e_i) = f(e_i | E_{i-1})$.
- Then, we have generated a point x (a vertex, no less) in B_f such that $x(A) = f(A)$.
- Thus, for any A, we have

$$B_f \cap \left\{ x \in \mathbb{R}^E : x(A) = f(A) \right\} \neq \emptyset \tag{15.2}$$

- In words, B_f intersects all "multi-axis congruent" hyperplanes within R^E of the form $\left\{ x \in \mathbb{R}^E : x(A) = f(A) \right\}$ for all $A \subseteq E$.

B_f dominates P_f

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

Theorem 15.5.2

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

Proof.

...

B_f dominates P_f

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

Theorem 15.5.2

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

Proof.

- We construct the y algorithmically: initially set $y \leftarrow x$.

...

B_f dominates P_f

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

Theorem 15.5.2

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

Proof.

- We construct the y algorithmically: initially set $y \leftarrow x$.
- $y \in P_f$, T is tight for y so $y(T) = f(T)$.

...

B_f dominates P_f

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

Theorem 15.5.2

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

Proof.

- We construct the y algorithmically: initially set $y \leftarrow x$.
- $y \in P_f$, T is tight for y so $y(T) = f(T)$.
- Recall saturation capacity: for $y \in P_f$, $\hat{c}(y; e) = \min\{f(A) - y(A) | \forall A \ni e\} = \max\{\alpha : \alpha \in \mathbb{R}, y + \alpha \mathbf{1}_e \in P_f\}$

...

B_f dominates P_f

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

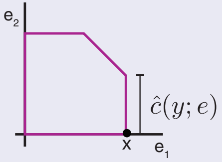
Theorem 15.5.2

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

Proof.

- We construct the y algorithmically: initially set $y \leftarrow x$.
- $y \in P_f$, T is tight for y so $y(T) = f(T)$.
- Recall saturation capacity: for $y \in P_f$, $\hat{c}(y;e) = \min\{f(A) - y(A) | \forall A \ni e\} = \max\{\alpha : \alpha \in \mathbb{R}, y + \alpha \mathbf{1}_e \in P_f\}$
- Consider following algorithm:

```
T' ← T ;
for e ∈ E \ T do
    y ← y + c(y; e)1_e ; T' ← T' ∪ {e};
```



...

B_f dominates P_f

... proof of Thm. 15.5.2 cont.

- Each step maintains feasibility: consider one step adding e to T' — for $e \notin T'$, feasibility requires $y(T' + e) = y(T') + y(e) \leq f(T' + e)$, or $y(e) \leq f(T' + e) - y(T')$.

B_f dominates P_f

. . . proof of Thm. 15.5.2 cont.

- Each step maintains feasibility: consider one step adding e to T' — for $e \notin T'$, feasibility requires $y(T' + e) = y(T') + y(e) \leq f(T' + e)$, or $y(e) \leq f(T' + e) - y(T') = y(e) + f(T' + e) - y(T' + e)$.

B_f dominates P_f

. . . proof of Thm. 15.5.2 cont.

- Each step maintains feasibility: consider one step adding e to T' — for $e \notin T'$, feasibility requires $y(T' + e) = y(T') + y(e) \leq f(T' + e)$, or $y(e) \leq f(T' + e) - y(T') = y(e) + f(T' + e) - y(T' + e)$.
- We set $y(e) \leftarrow y(e) + \hat{c}(y; e) \leq y(e) + f(T' + e) - y(T' + e)$.

B_f dominates P_f

... proof of Thm. 15.5.2 cont.

- Each step maintains feasibility: consider one step adding e to T' — for $e \notin T'$, feasibility requires $y(T' + e) = y(T') + y(e) \leq f(T' + e)$, or $y(e) \leq f(T' + e) - y(T') = y(e) + f(T' + e) - y(T' + e)$.
- We set $y(e) \leftarrow y(e) + \hat{c}(y; e) \leq y(e) + f(T' + e) - y(T' + e)$. Hence, after each step, $y \in P_f$ and $\hat{c}(y; e) \geq 0$. (also, consider r.h. version of $\hat{c}(y; e)$).

B_f dominates P_f

. . . proof of Thm. 15.5.2 cont.

- Each step maintains feasibility: consider one step adding e to T' — for $e \notin T'$, feasibility requires $y(T' + e) = y(T') + y(e) \leq f(T' + e)$, or $y(e) \leq f(T' + e) - y(T') = y(e) + f(T' + e) - y(T' + e)$.
- We set $y(e) \leftarrow y(e) + \hat{c}(y; e) \leq y(e) + f(T' + e) - y(T' + e)$. Hence, after each step, $y \in P_f$ and $\hat{c}(y; e) \geq 0$. (also, consider r.h. version of $\hat{c}(y; e)$).
- Also, only $y(e)$ for $e \notin T$ changed, final y has $y(e) = x(e)$ for $e \in T$.

B_f dominates P_f

... proof of Thm. 15.5.2 cont.

- Each step maintains feasibility: consider one step adding e to T' — for $e \notin T'$, feasibility requires $y(T' + e) = y(T') + y(e) \leq f(T' + e)$, or $y(e) \leq f(T' + e) - y(T') = y(e) + f(T' + e) - y(T' + e)$.
- We set $y(e) \leftarrow y(e) + \hat{c}(y; e) \leq y(e) + f(T' + e) - y(T' + e)$. Hence, after each step, $y \in P_f$ and $\hat{c}(y; e) \geq 0$. (also, consider r.h. version of $\hat{c}(y; e)$).
- Also, only $y(e)$ for $e \notin T$ changed, final y has $y(e) = x(e)$ for $e \in T$.
- Let $S_e \ni e$ be a set that achieves $c(y; e) = f(S_e) - y(S_e)$.

B_f dominates P_f

. . . proof of Thm. 15.5.2 cont.

- Each step maintains feasibility: consider one step adding e to T' — for $e \notin T'$, feasibility requires $y(T' + e) = y(T') + y(e) \leq f(T' + e)$, or $y(e) \leq f(T' + e) - y(T') = y(e) + f(T' + e) - y(T' + e)$.
- We set $y(e) \leftarrow y(e) + \hat{c}(y; e) \leq y(e) + f(T' + e) - y(T' + e)$. Hence, after each step, $y \in P_f$ and $\hat{c}(y; e) \geq 0$. (also, consider r.h. version of $\hat{c}(y; e)$).
- Also, only $y(e)$ for $e \notin T$ changed, final y has $y(e) = x(e)$ for $e \in T$.
- Let $S_e \ni e$ be a set that achieves $c(y; e) = f(S_e) - y(S_e)$.
- At iteration e, let $y'(e)$ (resp. $y(e)$) be new (resp. old) entry for e, then

$$y'(S_e) = y(S_e \setminus \{e\}) + y'(e) \tag{15.3}$$
$$= y(S_e \setminus \{e\}) + [y(e) + f(S_e) - y(S_e)] = f(S_e)$$

B_f dominates P_f

...proof of Thm. 15.5.2 cont.

- Each step maintains feasibility: consider one step adding e to T' — for $e \notin T'$, feasibility requires $y(T'+e) = y(T') + y(e) \leq f(T'+e)$, or $y(e) \leq f(T'+e) - y(T') = y(e) + f(T'+e) - y(T'+e)$.
- We set $y(e) \leftarrow y(e) + \hat{c}(y;e) \leq y(e) + f(T'+e) - y(T'+e)$. Hence, after each step, $y \in P_f$ and $\hat{c}(y;e) \geq 0$. (also, consider r.h. version of $\hat{c}(y;e)$).
- Also, only $y(e)$ for $e \notin T$ changed, final y has $y(e) = x(e)$ for $e \in T$.
- Let $S_e \ni e$ be a set that achieves $c(y;e) = f(S_e) - y(S_e)$.
- At iteration e, let $y'(e)$ (resp. $y(e)$) be new (resp. old) entry for e, then

$$\begin{aligned} y'(S_e) &= y(S_e \setminus \{e\}) + y'(e) && (15.3) \\ &= y(S_e \setminus \{e\}) + [y(e) + f(S_e) - y(S_e)] = f(S_e) \end{aligned}$$

So, S_e is tight for y'. It remains tight in further iterations since y doesn't decrease and it stays within P_f.

B_f dominates P_f

...proof of Thm. 15.5.2 cont.

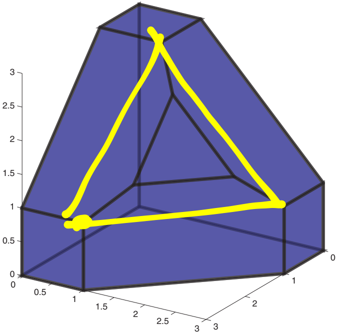
- Each step maintains feasibility: consider one step adding e to T' — for $e \notin T'$, feasibility requires $y(T'+e) = y(T') + y(e) \leq f(T'+e)$, or $y(e) \leq f(T'+e) - y(T') = y(e) + f(T'+e) - y(T'+e)$.
- We set $y(e) \leftarrow y(e) + \hat{c}(y;e) \leq y(e) + f(T'+e) - y(T'+e)$. Hence, after each step, $y \in P_f$ and $\hat{c}(y;e) \geq 0$. (also, consider r.h. version of $\hat{c}(y;e)$).
- Also, only $y(e)$ for $e \notin T$ changed, final y has $y(e) = x(e)$ for $e \in T$.
- Let $S_e \ni e$ be a set that achieves $c(y;e) = f(S_e) - y(S_e)$.
- At iteration e, let $y'(e)$ (resp. $y(e)$) be new (resp. old) entry for e, then

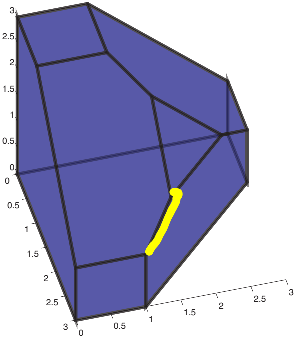
$$\begin{aligned} y'(S_e) &= y(S_e \setminus \{e\}) + y'(e) \qquad (15.3) \\ &= y(S_e \setminus \{e\}) + [y(e) + f(S_e) - y(S_e)] = f(S_e) \end{aligned}$$

  So, S_e is tight for y'. It remains tight in further iterations since y doesn't decrease and it stays within P_f.
- Also, $E = T \cup \bigcup_{e \notin T} S_e$ is also tight, meaning the final y has $y \in B_f$.

Polytope example 1

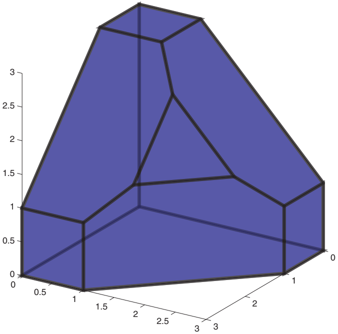
- Observe: P_f (at two views):

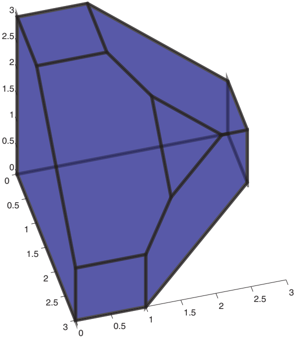




Polytope example 1

- Observe: P_f (at two views):

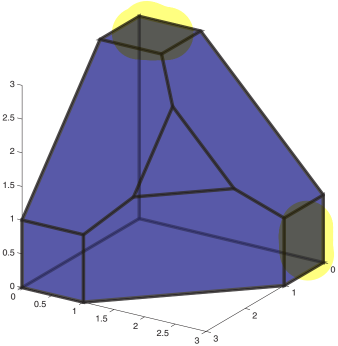


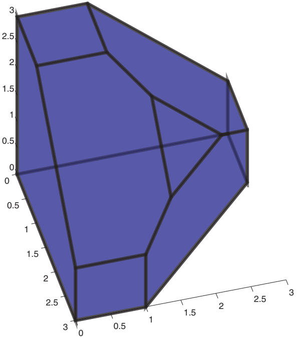


- Is this a polymatroidal polytope?

Polytope example 1

- Observe: P_f (at two views):

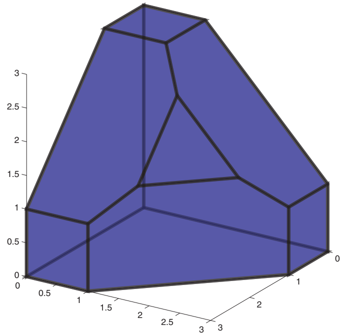


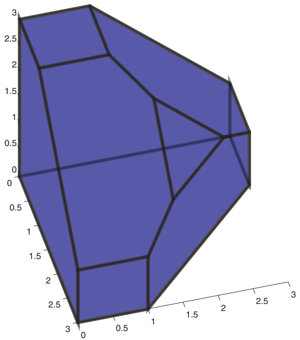


- Is this a polymatroidal polytope?
- No, "B_f" doesn't intersect sets of the form $\{x : x(e) = f(e)\}$ for $e \in E$.

Polytope example 1

- Observe: P_f (at two views):

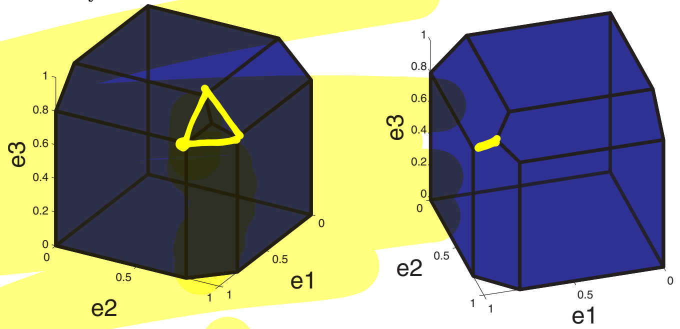




- Is this a polymatroidal polytope?
- No, "B_f" doesn't intersect sets of the form $\{x : x(e) = f(e)\}$ for $e \in E$.
- This was generated using function $g(0) = 0$, $g(1) = 3$, $g(2) = 4$, and $g(3) = 5.5$. Then $f(S) = g(|S|)$ is not submodular since (e.g.) $f(\{e_1, e_3\}) + f(\{e_1, e_2\}) = 4 + 4 = 8$ but

Polytope example 2

- Observe: P_f (at two views):

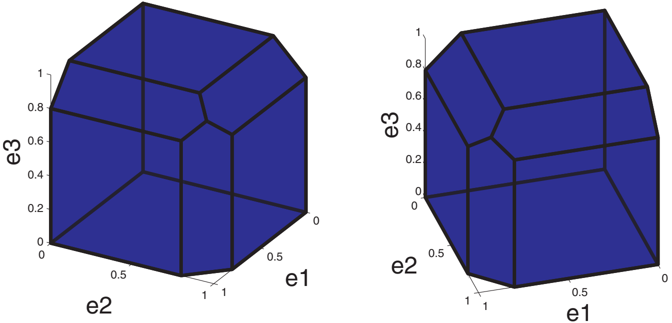


$x : x(e_1) = f(e_1)$

$x : x(e_1, e_2) = f(e_1, e_2)$

Polytope example 2

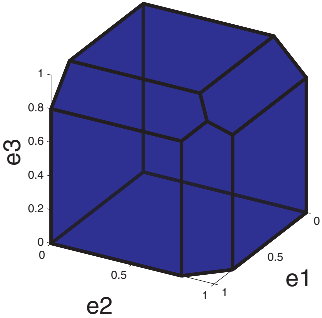
- Observe: P_f (at two views):

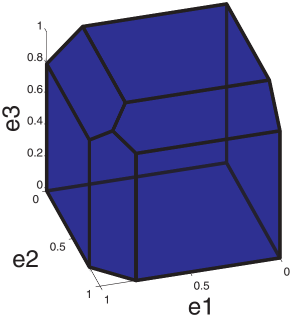


- Is this a polymatroidal polytope?

Polytope example 2

- Observe: P_f (at two views):

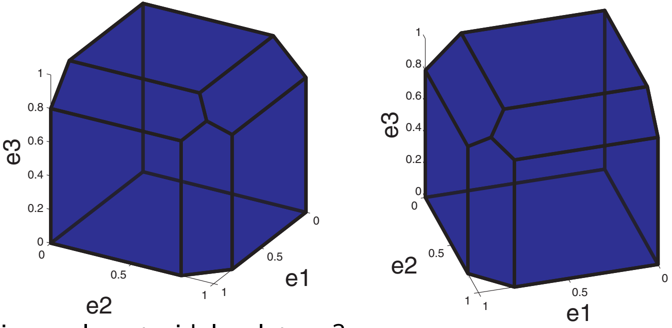




- Is this a polymatroidal polytope?
- No, "B_f" (which would be a single point in this case) doesn't intersect sets of the form $\{x : x(e) = f(e)\}$ for $e \in E$.

Polytope example 2

- Observe: P_f (at two views):



- Is this a polymatroidal polytope?
- No, "B_f" (which would be a single point in this case) doesn't intersect sets of the form $\{x : x(e) = f(e)\}$ for $e \in E$.
- This was generated using function $g(0) = 0$, $g(1) = 1$, $g(2) = 1.8$, and $g(3) = 3$. Then $f(S) = g(|S|)$ is not submodular since (e.g.) $f(\{e_1, e_3\}) + f(\{e_1, e_2\}) = 1.8 + 1.8 = 3.6$ but $f(\{e_1, e_2, e_3\}) + f(\{e_1\}) = 3 + 1 = 4$.

Review

The next slide is review from lecture 13.

Saturation Capacity

- The max is achieved when

$$\alpha = \hat{c}(x;e) \stackrel{\text{def}}{=} \min\{f(A) - x(A), \forall A \supseteq \{e\}\} \tag{15.22}$$

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

Matroids and Exchange

- Recall, matroids have a number of "exchange" properties.

Matroids and Exchange

- Recall, matroids have a number of "exchange" properties.
- Also, recall that given a matroid $\mathcal{M} = (E, \mathcal{I})$, if $I \in \mathcal{I}$ is independent, and $e \in \text{span}(I)$, and $e' \in C(I, e)$ where $C(I, e)$ is the fundamental circuit created when adding e to I, then we have:

$$I + e - e' \in \mathcal{I} \tag{15.4}$$

Matroids and Exchange

- Recall, matroids have a number of "exchange" properties.
- Also, recall that given a matroid $\mathcal{M} = (E, \mathcal{I})$, if $I \in \mathcal{I}$ is independent, and $e \in \text{span}(I)$, and $e' \in C(I, e)$ where $C(I, e)$ is the fundamental circuit created when adding e to I, then we have:

$$I + e - e' \in \mathcal{I} \tag{15.4}$$

- Note, this holds for any $e' \in C(I, e)$.

Matroids and Exchange

- Recall, matroids have a number of "exchange" properties.
- Also, recall that given a matroid $\mathcal{M} = (E, \mathcal{I})$, if $I \in \mathcal{I}$ is independent, and $e \in \text{span}(I)$, and $e' \in C(I, e)$ where $C(I, e)$ is the fundamental circuit created when adding e to I, then we have:

$$I + e - e' \in \mathcal{I} \tag{15.4}$$

- Note, this holds for any $e' \in C(I, e)$.
- Since $\text{dep}(x, e)$ generalizes the fundamental circuit of a matroid to polymatroids, we saw (last lecture) that this a property exists for polymatroids as well.

Matroids and Exchange

- Recall, matroids have a number of "exchange" properties.
- Also, recall that given a matroid $\mathcal{M} = (E, \mathcal{I})$, if $I \in \mathcal{I}$ is independent, and $e \in \text{span}(I)$, and $e' \in C(I, e)$ where $C(I, e)$ is the fundamental circuit created when adding e to I, then we have:

$$I + e - e' \in \mathcal{I} \tag{15.4}$$

- Note, this holds for any $e' \in C(I, e)$.
- Since $\text{dep}(x, e)$ generalizes the fundamental circuit of a matroid to polymatroids, we saw (last lecture) that this a property exists for polymatroids as well.
- As there is saturation capacity for elements that are not saturated, is there is a corresponding concept for degree of polymatroidal exchange?

Matroids and Exchange

- Recall, matroids have a number of "exchange" properties.
- Also, recall that given a matroid $\mathcal{M} = (E, \mathcal{I})$, if $I \in \mathcal{I}$ is independent, and $e \in \text{span}(I)$, and $e' \in C(I, e)$ where $C(I, e)$ is the fundamental circuit created when adding e to I, then we have:

$$I + e - e' \in \mathcal{I} \tag{15.4}$$

- Note, this holds for any $e' \in C(I, e)$.
- Since $\text{dep}(x, e)$ generalizes the fundamental circuit of a matroid to polymatroids, we saw (last lecture) that this a property exists for polymatroids as well.
- As there is saturation capacity for elements that are not saturated, is there is a corresponding concept for degree of polymatroidal exchange?
- Yes, and it is called the "exchange capacity"

Exchange Capacity

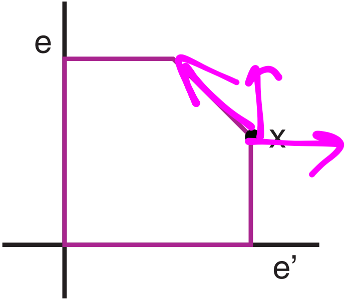
- Consider $x \in P_f$, $e \in \text{sat}(x)$ and $e' \in \text{dep}(x, e) \setminus \{e\}$

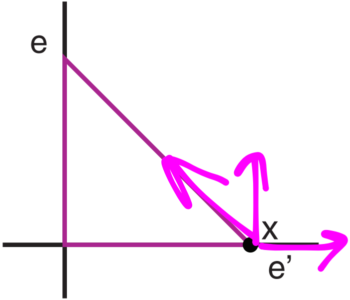
Exchange Capacity

- Consider $x \in P_f$, $e \in \text{sat}(x)$ and $e' \in \text{dep}(x, e) \setminus \{e\}$
- Thus, for any $\alpha > 0$, we have $x + \alpha \mathbf{1}_a \notin P_f$ for either $a = e$ or $a = e'$, since $\text{dep}(x, e) \subseteq \text{sat}(x)$.

Exchange Capacity

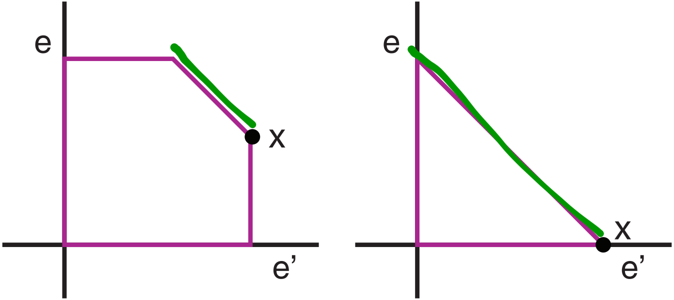
- Consider $x \in P_f$, $e \in \text{sat}(x)$ and $e' \in \text{dep}(x,e) \setminus \{e\}$
- Thus, for any $\alpha > 0$, we have $x + \alpha \mathbf{1}_a \notin P_f$ for either $a = e$ or $a = e'$, since $\text{dep}(x,e) \subseteq \text{sat}(x)$.
- Examples:





Exchange Capacity

- Consider $x \in P_f$, $e \in \text{sat}(x)$ and $e' \in \text{dep}(x, e) \setminus \{e\}$
- Thus, for any $\alpha > 0$, we have $x + \alpha \mathbf{1}_a \notin P_f$ for either $a = e$ or $a = e'$, since $\text{dep}(x, e) \subseteq \text{sat}(x)$.
- Examples:



- How much can we move in positive e direction if we simultaneously move in negative e' direction?

Exchange Capacity

- $x \in P_f$, $e \in \text{sat}(x)$ and $e' \in \text{dep}(x, e) \setminus \{e\}$, consider

$$\max \{\alpha : \alpha \in \mathbb{R}, x + \alpha(\mathbf{1}_e - \mathbf{1}_{e'}) \in P_f\} \tag{15.5}$$

Exchange Capacity

- $x \in P_f$, $e \in \mathrm{sat}(x)$ and $e' \in \mathrm{dep}(x,e) \setminus \{e\}$, consider
$$\max\{\alpha : \alpha \in \mathbb{R}, x + \alpha(\mathbf{1}_e - \mathbf{1}_{e'}) \in P_f\} \tag{15.5}$$
- Identical to:
$$\max\{\alpha : \alpha \in \mathbb{R}, (x + \alpha(\mathbf{1}_e - \mathbf{1}_{e'}))(A) \le f(A), \forall A\} \tag{15.6}$$

(A)

Exchange Capacity

- $x \in P_f$, $e \in \text{sat}(x)$ and $e' \in \text{dep}(x,e) \setminus \{e\}$, consider
$$\max\{\alpha : \alpha \in \mathbb{R}, x + \alpha(\mathbf{1}_e - \mathbf{1}_{e'}) \in P_f\} \tag{15.5}$$
- Identical to:
$$\max\{\alpha : \alpha \in \mathbb{R}, (x + \alpha(\mathbf{1}_e - \mathbf{1}_{e'}))(A) \leq f(A), \forall A\} \tag{15.6}$$
- If both $e, e' \in A$ (or neither), then $\alpha(\mathbf{1}_e - \mathbf{1}_{e'})(A) = 0$ for any α, so to make this meaningful, we take $A : e' \notin A \ni e$.

Exchange Capacity

- $x \in P_f$, $e \in \text{sat}(x)$ and $e' \in \text{dep}(x,e) \setminus \{e\}$, consider

$$\max\{\alpha : \alpha \in \mathbb{R}, x + \alpha(\mathbf{1}_e - \mathbf{1}_{e'}) \in P_f\} \tag{15.5}$$

- Identical to:

$$\max\{\alpha : \alpha \in \mathbb{R}, (x + \alpha(\mathbf{1}_e - \mathbf{1}_{e'}))(A) \leq f(A), \forall A\} \tag{15.6}$$

- If both $e, e' \in A$ (or neither), then $\alpha(\mathbf{1}_e - \mathbf{1}_{e'})(A) = 0$ for any α, so to make this meaningful, we take $A : e' \notin A \ni e$.
- thus identical to

$$\max\{\alpha : \alpha \in \mathbb{R}, (x + \alpha(\mathbf{1}_e - \mathbf{1}_{e'}))(A) \leq f(A), \forall A \ni e, e' \notin A\} \tag{15.7}$$

Exchange Capacity

- $x \in P_f$, $e \in \text{sat}(x)$ and $e' \in \text{dep}(x, e) \setminus \{e\}$, consider
$$\max \{\alpha : \alpha \in \mathbb{R}, x + \alpha(\mathbf{1}_e - \mathbf{1}_{e'}) \in P_f\} \tag{15.5}$$
- Identical to:
$$\max \{\alpha : \alpha \in \mathbb{R}, (x + \alpha(\mathbf{1}_e - \mathbf{1}_{e'}))(A) \le f(A), \forall A\} \tag{15.6}$$
- If both $e, e' \in A$ (or neither), then $\alpha(\mathbf{1}_e - \mathbf{1}_{e'})(A) = 0$ for any α, so to make this meaningful, we take $A : e' \notin A \ni e$.
- thus identical to
$$\max \{\alpha : \alpha \in \mathbb{R}, (x + \alpha(\mathbf{1}_e - \mathbf{1}_{e'}))(A) \le f(A), \forall A \ni e, e' \notin A\} \tag{15.7}$$
- Which is identical to:
$$\max \{\alpha : \alpha \in \mathbb{R}, \alpha(\mathbf{1}_e - \mathbf{1}_{e'}))(A) \le f(A) - x(A), \forall A \supseteq \{e\}, e' \notin A\} \tag{15.8}$$

Exchange Capacity

- In such case, we get $\mathbf{1}_{e'}(A) = 0$, thus above identical to

$$\max \left\{ \alpha : \alpha \in \mathbb{R}, \alpha \mathbf{1}_e(A) \leq f(A) - x(A), \forall A \supseteq \{e\}, e' \notin A \right\} \quad (15.9)$$

Exchange Capacity

- In such case, we get $\mathbf{1}_{e'}(A) = 0$, thus above identical to

$$\max \left\{ \alpha : \alpha \in \mathbb{R}, \alpha \mathbf{1}_e(A) \leq f(A) - x(A), \forall A \supseteq \{e\}, e' \notin A \right\} \quad (15.9)$$

- Restating, we've got

$$\max \left\{ \alpha : \alpha \in \mathbb{R}, \alpha \leq f(A) - x(A), \forall A \supseteq \{e\}, e' \notin A \right\} \quad (15.10)$$

Exchange Capacity

- In such case, we get $\mathbf{1}_{e'}(A) = 0$, thus above identical to

$$\max\left\{\alpha : \alpha \in \mathbb{R}, \alpha \mathbf{1}_e(A) \leq f(A) - x(A), \forall A \supseteq \{e\}, e' \notin A\right\} \quad (15.9)$$

- Restating, we've got

$$\max\left\{\alpha : \alpha \in \mathbb{R}, \alpha \leq f(A) - x(A), \forall A \supseteq \{e\}, e' \notin A\right\} \quad (15.10)$$

- This max is achieved when

$$\alpha = \hat{c}(x; e, e') \stackrel{\text{def}}{=} \min\left\{f(A) - x(A), \forall A \supseteq \{e\}, e' \notin A\right\} \quad (15.11)$$

Exchange Capacity

- In such case, we get $\mathbf{1}_{e'}(A) = 0$, thus above identical to

$$\max\left\{\alpha : \alpha \in \mathbb{R}, \alpha \mathbf{1}_e(A) \leq f(A) - x(A), \forall A \supseteq \{e\}, e' \notin A\right\} \quad (15.9)$$

- Restating, we've got

$$\max\left\{\alpha : \alpha \in \mathbb{R}, \alpha \leq f(A) - x(A), \forall A \supseteq \{e\}, e' \notin A\right\} \quad (15.10)$$

- This max is achieved when

$$\alpha = \hat{c}(x; e, e') \stackrel{\text{def}}{=} \min\left\{f(A) - x(A), \forall A \supseteq \{e\}, e' \notin A\right\} \quad (15.11)$$

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

Exchange Capacity

- In such case, we get $\mathbf{1}_{e'}(A) = 0$, thus above identical to

$$\max\left\{\alpha : \alpha \in \mathbb{R}, \alpha \mathbf{1}_e(A) \le f(A) - x(A), \forall A \supseteq \{e\}, e' \notin A\right\} \quad (15.9)$$

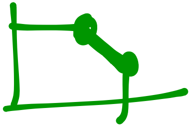
- Restating, we've got

$$\max\left\{\alpha : \alpha \in \mathbb{R}, \alpha \le f(A) - x(A), \forall A \supseteq \{e\}, e' \notin A\right\} \quad (15.10)$$

- This max is achieved when

$$\alpha = \hat{c}(x; e, e') \stackrel{\text{def}}{=} \min\left\{f(A) - x(A), \forall A \supseteq \{e\}, e' \notin A\right\} \quad (15.11)$$

- $\hat{c}(x; e, e')$ is known as the exchange capacity associated with $x \in P_f$ and e.
- For any α with $0 \le \alpha \le \hat{c}(x; e, e')$, we have that $x + \alpha(\mathbf{1}_e - \mathbf{1}_{e'}) \in P_f$.



Exchange Capacity

- In such case, we get $\mathbf{1}_{e'}(A) = 0$, thus above identical to

$$\max\left\{\alpha : \alpha \in \mathbb{R}, \alpha \mathbf{1}_e(A) \leq f(A) - x(A), \forall A \supseteq \{e\}, e' \notin A\right\} \quad (15.9)$$

- Restating, we've got

$$\max\left\{\alpha : \alpha \in \mathbb{R}, \alpha \leq f(A) - x(A), \forall A \supseteq \{e\}, e' \notin A\right\} \quad (15.10)$$

- This max is achieved when

$$\alpha = \hat{c}(x; e, e') \stackrel{\text{def}}{=} \min\left\{f(A) - x(A), \forall A \supseteq \{e\}, e' \notin A\right\} \quad (15.11)$$

- $\hat{c}(x; e, e')$ is known as the exchange capacity associated with $x \in P_f$ and e.
- For any α with $0 \leq \alpha \leq \hat{c}(x; e, e')$, we have that $x + \alpha(\mathbf{1}_e - \mathbf{1}_{e'}) \in P_f$.
- As we will see, if e and e' are successive in an order that generates extreme point x, then we get a "neighbor" extreme point via $x' = x + \hat{c}(x; e, e')(\mathbf{1}_e - \mathbf{1}_{e'})$.

Exchange Capacity

- In such case, we get $\mathbf{1}_{e'}(A) = 0$, thus above identical to

$$\max\left\{\alpha : \alpha \in \mathbb{R}, \alpha \mathbf{1}_e(A) \leq f(A) - x(A), \forall A \supseteq \{e\}, e' \notin A\right\} \quad (15.9)$$

- Restating, we've got

$$\max\left\{\alpha : \alpha \in \mathbb{R}, \alpha \leq f(A) - x(A), \forall A \supseteq \{e\}, e' \notin A\right\} \quad (15.10)$$

- This max is achieved when

$$\alpha = \hat{c}(x; e, e') \stackrel{\text{def}}{=} \min\left\{f(A) - x(A), \forall A \supseteq \{e\}, e' \notin A\right\} \quad (15.11)$$

- $\hat{c}(x; e, e')$ is known as the exchange capacity associated with $x \in P_f$ and e.
- For any α with $0 \leq \alpha \leq \hat{c}(x; e, e')$, we have that $x + \alpha(\mathbf{1}_e - \mathbf{1}_{e'}) \in P_f$.
- As we will see, if e and e' are successive in an order that generates extreme point x, then we get a "neighbor" extreme point via $x' = x + \hat{c}(x; e, e')(\mathbf{1}_e - \mathbf{1}_{e'})$.
- Note that Eqn. (15.11) is a form of SFM.

A polymatroid function's polyhedron is a polymatroid.

Theorem 15.7.1

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

$$\boxed{\textit{rank}(x) = \max\left(y(E) : y \leq x, y \in P_f\right) = \min\left(x(A) + f(E \setminus A) : A \subseteq E\right)} \quad (15.5)$$

$P_f = \{x \in \mathbb{R}^E : x(A) \leq f(A)\}$

If we take x to be zero, we get:

Corollary 15.7.2

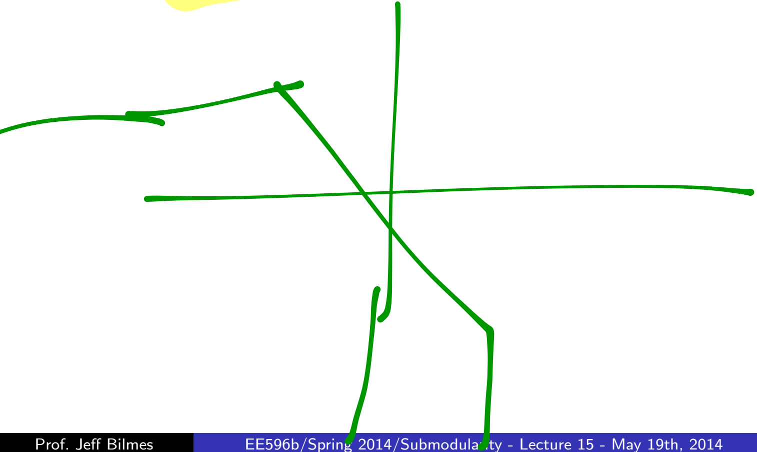
Let f be a submodular function defined on subsets of E. $x \in \mathbb{R}^E$, we have:

$$\textit{rank}(0) = \max\left(y(E) : y \leq 0, y \in P_f\right) = \min\left(f(A) : A \subseteq E\right) \quad (15.6)$$

Min-Norm Point: Definition

- Restating what we saw before, we have:

$$\max\{y(E) \mid y \in P_f, y \leq 0\} = \min\{f(X) \mid X \subseteq V\} \tag{15.12}$$



Min-Norm Point: Definition

- Restating what we saw before, we have:

$$\max\{y(E) | y \in P_f, y \leq 0\} = \min\{f(X) | X \subseteq V\} \tag{15.12}$$

- Consider the optimization:

$$\text{minimize} \quad \|x\|_2^2 = \sum_e (x(e))^2 \tag{15.13a}$$

$$\text{subject to} \quad x \in B_f \tag{15.13b}$$

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

Min-Norm Point: Definition

- Restating what we saw before, we have:

$$\max\{y(E)|y \in P_f, y \leq 0\} = \min\{f(X)|X \subseteq V\} \tag{15.12}$$

- Consider the optimization:

$$\begin{aligned} &\text{minimize} && \|x\|_2^2 && (15.13a)\\ &\text{subject to} && x \in B_f && (15.13b) \end{aligned}$$

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

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

Min-Norm Point: Definition

- Restating what we saw before, we have:

$$\max\{y(E) | y \in P_f, y \leq 0\} = \min\{f(X) | X \subseteq V\} \tag{15.12}$$

- Consider the optimization:

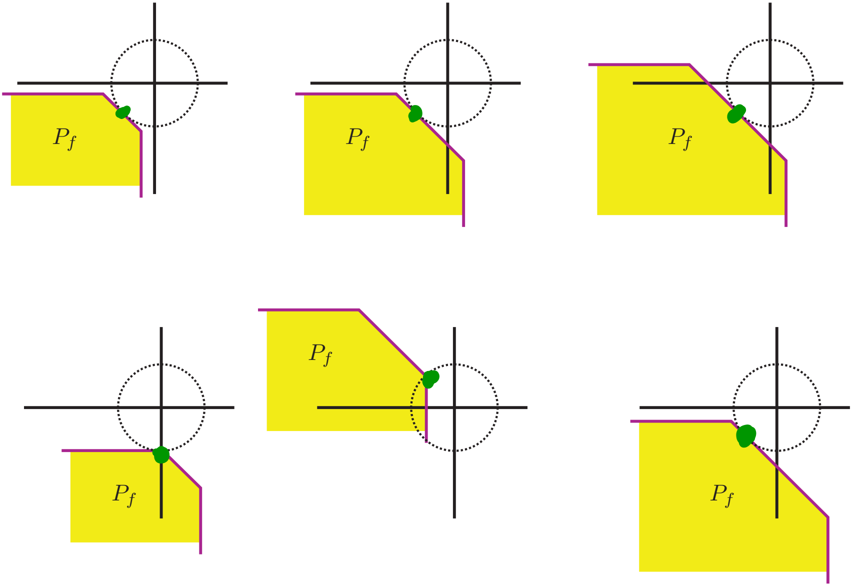
$$\text{minimize} \quad \|x\|_2^2 \tag{15.13a}$$

$$\text{subject to} \quad x \in B_f \tag{15.13b}$$

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

- Note, x^* is the unique optimal solution since we have a strictly convex objective over a set of convex constraints.
- x^* is called the minimum norm point of the base polytope.

Min-Norm Point: Examples



Min-Norm Point and Submodular Function Minimization

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

$$y^* = x^* \wedge 0 = (\min(x^*(e), 0) | e \in E) \tag{15.14}$$
$$A_- = \{e : x^*(e) < 0\} \tag{15.15}$$
$$A_0 = \{e : x^*(e) \leq 0\} \tag{15.16}$$

Min-Norm Point and Submodular Function Minimization

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

$$y^* = x^* \wedge 0 = (\min(x^*(e), 0) | e \in E) \tag{15.14}$$

$$A_- = \{e : x^*(e) < 0\} \tag{15.15}$$

$$A_0 = \{e : x^*(e) \leq 0\} \tag{15.16}$$

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

Min-Norm Point and Submodular Function Minimization

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

$$y^* = x^* \wedge 0 = (\min(x^*(e), 0) | e \in E) \tag{15.14}$$

$$A_- = \{e : x^*(e) < 0\} \tag{15.15}$$

$$A_0 = \{e : x^*(e) \leq 0\} \tag{15.16}$$

- It turns out, these quantities will solve the submodular function minimization problem, as we now show.
- The proof is nice since it uses the tools we've been recently developing.

Min-Norm Point and SFM

Theorem 15.7.1

Let y^*, A_-, and A_0 be as given. Then y^* is a maximizer of the l.h.s. of Eqn. (15.12)*. Moreover, A_- is the unique minimal minimizer of f and A_0 is the unique maximal minimizer of f.*

Proof.

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

...

Min-Norm Point and SFM

Theorem 15.7.1

Let y^*, A_-, and A_0 be as given. Then y^* is a maximizer of the l.h.s. of Eqn. (15.12). Moreover, A_- is the unique minimal minimizer of f and A_0 is the unique maximal minimizer of f.

Proof.

- First note, since $x^* \in B_f$, we have $x^*(E) = f(E)$, meaning $\mathrm{sat}(x^*) = E$. Thus, we can consider any $e \in E$ within $\mathrm{dep}(x^*, e)$.
- Consider any pair (e, e') with $e' \in \mathrm{dep}(x^*, e)$ and $e \in A_-$. Then $x^*(e) < 0$, and $\exists \alpha > 0$ s.t. $x^* + \alpha \mathbf{1}_e - \alpha \mathbf{1}_{e'} \in P_f$.

. . .

Min-Norm Point and SFM

Theorem 15.7.1

Let y^*, A_-, and A_0 be as given. Then y^* is a maximizer of the l.h.s. of Eqn. (15.12). Moreover, A_- is the unique minimal minimizer of f and A_0 is the unique maximal minimizer of f.

Proof.

- First note, since $x^* \in B_f$, we have $x^*(E) = f(E)$, meaning $\mathrm{sat}(x^*) = E$. Thus, we can consider any $e \in E$ within $\mathrm{dep}(x^*, e)$.
- Consider any pair (e, e') with $e' \in \mathrm{dep}(x^*, e)$ and $e \in A_-$. Then $x^*(e) < 0$, and $\exists \alpha > 0$ s.t. $x^* + \alpha \mathbf{1}_e - \alpha \mathbf{1}_{e'} \in P_f$.
- We have $x^*(E) = f(E)$ and x^* is minimum in l2 sense. We have $(x^* + \alpha \mathbf{1}_e - \alpha \mathbf{1}_{e'}) \in P_f$, and in fact

$$(x^* + \alpha \mathbf{1}_e - \alpha \mathbf{1}_{e'})(E) = x^*(E) + \alpha - \alpha = f(E) \tag{15.17}$$

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

...

Min-Norm Point and SFM

. . . proof of Thm. 15.7.1 cont.

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

. . .

Min-Norm Point and SFM

. . . proof of Thm. 15.7.1 cont.

- Then $(x^* + \alpha \mathbf{1}_e - \alpha \mathbf{1}_{e'})(E)$
  $= x^*(E \setminus \{e, e'\}) + \underbrace{(x^*(e) + \alpha)}_{x^*_{\text{new}}(e)} + \underbrace{(x^*(e') - \alpha)}_{x^*_{\text{new}}(e')} = f(E).$
- Minimality of $x^* \in B_f$ in l2 sense requires that, with such an $\alpha > 0$,
  $\left(x^*(e)\right)^2 + \left(x^*(e')\right)^2 < \left(x^*_{\text{new}}(e)\right)^2 + \left(x^*_{\text{new}}(e')\right)^2$

. . .

Min-Norm Point and SFM

. . . proof of Thm. 15.7.1 cont.

- Then $(x^* + \alpha \mathbf{1}_e - \alpha \mathbf{1}_{e'})(E)$
  $= x^*(E \setminus \{e, e'\}) + \underbrace{(x^*(e) + \alpha)}_{x^*_{\text{new}}(e)} + \underbrace{(x^*(e') - \alpha)}_{x^*_{\text{new}}(e')} = f(E)$.
- Minimality of $x^* \in B_f$ in l2 sense requires that, with such an $\alpha > 0$,
  $\Big(x^*(e)\Big)^2 + \Big(x^*(e')\Big)^2 < \Big(x^*_{\text{new}}(e)\Big)^2 + \Big(x^*_{\text{new}}(e')\Big)^2$
- Given that $e \in A_-$, $x^*(e) < 0$. Thus, if $x^*(e') > 0$, we could have $(x^*(e) + \alpha)^2 + (x^*(e') - \alpha)^2 < (x^*(e))^2 + (x^*(e'))^2$, contradicting the optimality of x^*.

. . .

Min-Norm Point and SFM

. . . proof of Thm. 15.7.1 cont.

- Then $(x^* + \alpha \mathbf{1}_e - \alpha \mathbf{1}_{e'})(E)$
$= x^*(E \setminus \{e, e'\}) + \underbrace{(x^*(e) + \alpha)}_{x^*_{\text{new}}(e)} + \underbrace{(x^*(e') - \alpha)}_{x^*_{\text{new}}(e')} = f(E).$
- Minimality of $x^* \in B_f$ in l2 sense requires that, with such an $\alpha > 0$,
$\Big(x^*(e)\Big)^2 + \Big(x^*(e')\Big)^2 < \Big(x^*_{\text{new}}(e)\Big)^2 + \Big(x^*_{\text{new}}(e')\Big)^2$
- Given that $e \in A_-$, $x^*(e) < 0$. Thus, if $x^*(e') > 0$, we could have $(x^*(e) + \alpha)^2 + (x^*(e') - \alpha)^2 < (x^*(e))^2 + (x^*(e'))^2$, contradicting the optimality of x^*.
- If $x^*(e') = 0$, we would have $(x^*(e) + \alpha)^2 + (\alpha)^2 < (x^*(e))^2$, for any $0 < \alpha < |x^*(e)|$, again contradicting the optimality of x^*.

. . .

Min-Norm Point and SFM

. . . proof of Thm. 15.7.1 cont.

- Then $(x^* + \alpha \mathbf{1}_e - \alpha \mathbf{1}_{e'})(E)$ $= x^*(E \setminus \{e, e'\}) + \underbrace{(x^*(e) + \alpha)}_{x^*_{\text{new}}(e)} + \underbrace{(x^*(e') - \alpha)}_{x^*_{\text{new}}(e')} = f(E)$.
- Minimality of $x^* \in B_f$ in l2 sense requires that, with such an $\alpha > 0$, $\left(x^*(e)\right)^2 + \left(x^*(e')\right)^2 < \left(x^*_{\text{new}}(e)\right)^2 + \left(x^*_{\text{new}}(e')\right)^2$
- Given that $e \in A_-$, $x^*(e) < 0$. Thus, if $x^*(e') > 0$, we could have $(x^*(e) + \alpha)^2 + (x^*(e') - \alpha)^2 < (x^*(e))^2 + (x^*(e'))^2$, contradicting the optimality of x^*.
- If $x^*(e') = 0$, we would have $(x^*(e) + \alpha)^2 + (\alpha)^2 < (x^*(e))^2$, for any $0 < \alpha < |x^*(e)|$, again contradicting the optimality of x^*.
- Thus, we must have $x^*(e') < 0$ (strict negativity).

. . .

Min-Norm Point and SFM

. . . proof of Thm. 15.7.1 cont.

- Thus, for a pair (e, e') with $e' \in \mathrm{dep}(x^*, e)$ and $e \in A_-$, we have $x(e') < 0$ and hence $e' \in A_-$.

. . .

Min-Norm Point and SFM

. . . proof of Thm. 15.7.1 cont.

- Thus, for a pair (e, e') with $e' \in \text{dep}(x^*, e)$ and $e \in A_-$, we have $x(e') < 0$ and hence $e' \in A_-$.
- Hence, $\forall e \in A_-$, we have $\text{dep}(x^*, e) \subseteq A_-$.

. . .

Min-Norm Point and SFM

. . . proof of Thm. 15.7.1 cont.

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

. . .

Min-Norm Point and SFM

. . . proof of Thm. 15.7.1 cont.

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

Min-Norm Point and SFM

. . . proof of Thm. 15.7.1 cont.

- Therefore, we have $\cup_{e \in A_-} \operatorname{dep}(x^*, e) = A_-$ and $\cup_{e \in A_0} \operatorname{dep}(x^*, e) = A_0$
- Ie., $\{\operatorname{dep}(x^*, e)\}_{e \in A_-}$ is cover for A_-, as is $\{\operatorname{dep}(x^*, e)\}_{e \in A_0}$ for A_0.

Min-Norm Point and SFM

. . . proof of Thm. 15.7.1 cont.

- Therefore, we have $\cup_{e \in A_-} \text{dep}(x^*, e) = A_-$ and $\cup_{e \in A_0} \text{dep}(x^*, e) = A_0$
- Ie., $\{\text{dep}(x^*, e)\}_{e \in A_-}$ is cover for A_-, as is $\{\text{dep}(x^*, e)\}_{e \in A_0}$ for A_0.
- $\text{dep}(x^*, e)$ is minimal tight set containing e, meaning $x^*(\text{dep}(x^*, e)) = f(\text{dep}(x^*, e))$, and since tight sets are closed under union, we have that A_- and A_0 are also tight, meaning:

Min-Norm Point and SFM

. . . proof of Thm. 15.7.1 cont.

- Therefore, we have $\cup_{e \in A_-} \operatorname{dep}(x^*, e) = A_-$ and $\cup_{e \in A_0} \operatorname{dep}(x^*, e) = A_0$
- Ie., $\{\operatorname{dep}(x^*, e)\}_{e \in A_-}$ is cover for A_-, as is $\{\operatorname{dep}(x^*, e)\}_{e \in A_0}$ for A_0.
- $\operatorname{dep}(x^*, e)$ is minimal tight set containing e, meaning $x^*(\operatorname{dep}(x^*, e)) = f(\operatorname{dep}(x^*, e))$, and since tight sets are closed under union, we have that A_- and A_0 are also tight, meaning:

$$x^*(A_-) = f(A_-) \tag{15.18}$$

Min-Norm Point and SFM

. . . proof of Thm. 15.7.1 cont.

- Therefore, we have $\cup_{e \in A_-} \text{dep}(x^*, e) = A_-$ and $\cup_{e \in A_0} \text{dep}(x^*, e) = A_0$
- Ie., $\{\text{dep}(x^*, e)\}_{e \in A_-}$ is cover for A_-, as is $\{\text{dep}(x^*, e)\}_{e \in A_0}$ for A_0.
- $\text{dep}(x^*, e)$ is minimal tight set containing e, meaning $x^*(\text{dep}(x^*, e)) = f(\text{dep}(x^*, e))$, and since tight sets are closed under union, we have that A_- and A_0 are also tight, meaning:

$$x^*(A_-) = f(A_-) \tag{15.18}$$

$$x^*(A_0) = f(A_0) \tag{15.19}$$

Min-Norm Point and SFM

. . . proof of Thm. 15.7.1 cont.

- Therefore, we have $\cup_{e \in A_-} \text{dep}(x^*, e) = A_-$ and $\cup_{e \in A_0} \text{dep}(x^*, e) = A_0$
- Ie., $\{\text{dep}(x^*, e)\}_{e \in A_-}$ is cover for A_-, as is $\{\text{dep}(x^*, e)\}_{e \in A_0}$ for A_0.
- $\text{dep}(x^*, e)$ is minimal tight set containing e, meaning $x^*(\text{dep}(x^*, e)) = f(\text{dep}(x^*, e))$, and since tight sets are closed under union, we have that A_- and A_0 are also tight, meaning:

$$x^*(A_-) = f(A_-) \tag{15.18}$$

$$x^*(A_0) = f(A_0) \tag{15.19}$$

$$x^*(A_-) = x^*(A_0) = y^*(E) \tag{15.20}$$

Min-Norm Point and SFM

. . . proof of Thm. 15.7.1 cont.

- Therefore, we have $\cup_{e \in A_-} \operatorname{dep}(x^*, e) = A_-$ and $\cup_{e \in A_0} \operatorname{dep}(x^*, e) = A_0$
- Ie., $\{\operatorname{dep}(x^*, e)\}_{e \in A_-}$ is cover for A_-, as is $\{\operatorname{dep}(x^*, e)\}_{e \in A_0}$ for A_0.
- $\operatorname{dep}(x^*, e)$ is minimal tight set containing e, meaning $x^*(\operatorname{dep}(x^*, e)) = f(\operatorname{dep}(x^*, e))$, and since tight sets are closed under union, we have that A_- and A_0 are also tight, meaning:

$$x^*(A_-) = f(A_-) \tag{15.18}$$

$$x^*(A_0) = f(A_0) \tag{15.19}$$

$$x^*(A_-) = x^*(A_0) = y^*(E) \tag{15.20}$$

and therefore, all together we have

Min-Norm Point and SFM

. . . proof of Thm. 15.7.1 cont.

- Therefore, we have $\cup_{e \in A_-} \text{dep}(x^*, e) = A_-$ and $\cup_{e \in A_0} \text{dep}(x^*, e) = A_0$
- Ie., $\{\text{dep}(x^*, e)\}_{e \in A_-}$ is cover for A_-, as is $\{\text{dep}(x^*, e)\}_{e \in A_0}$ for A_0.
- $\text{dep}(x^*, e)$ is minimal tight set containing e, meaning $x^*(\text{dep}(x^*, e)) = f(\text{dep}(x^*, e))$, and since tight sets are closed under union, we have that A_- and A_0 are also tight, meaning:

$$x^*(A_-) = f(A_-) \tag{15.18}$$

$$x^*(A_0) = f(A_0) \tag{15.19}$$

$$x^*(A_-) = x^*(A_0) = y^*(E) \tag{15.20}$$

and therefore, all together we have

$$f(A_-) = f(A_0) = x^*(A_-) = x^*(A_0) = y^*(E) \tag{15.21}$$

Min-Norm Point and SFM

. . . proof of Thm. 15.7.1 cont.

- Now, y^* is feasible for the l.h.s. of Eqn. (15.12).

. . .

Min-Norm Point and SFM

. . . proof of Thm. 15.7.1 cont.

- Now, y^* is feasible for the l.h.s. of Eqn. (15.12). This follows since, we have $y^* = x^* \wedge 0 \leq 0$, and since $x^* \in B_f \subset P_f$, and $y^* \leq x^*$ and P_f is down-closed, we have that $y^* \in P_f$.

. . .

Min-Norm Point and SFM

. . . proof of Thm. 15.7.1 cont.

- Now, y^* is feasible for the l.h.s. of Eqn. (15.12). This follows since, we have $y^* = x^* \wedge 0 \leq 0$, and since $x^* \in B_f \subset P_f$, and $y^* \leq x^*$ and P_f is down-closed, we have that $y^* \in P_f$.
- Also, for any $y \in P_f$ with $y \leq 0$ and for any $X \subseteq E$, we have $y(E) \leq y(X) \leq f(X)$.

. . .

Min-Norm Point and SFM

. . . proof of Thm. 15.7.1 cont.

- Now, y^* is feasible for the l.h.s. of Eqn. (15.12). This follows since, we have $y^* = x^* \wedge 0 \leq 0$, and since $x^* \in B_f \subset P_f$, and $y^* \leq x^*$ and P_f is down-closed, we have that $y^* \in P_f$.
- Also, for any $y \in P_f$ with $y \leq 0$ and for any $X \subseteq E$, we have $y(E) \leq y(X) \leq f(X)$.
- Hence, we have found a feasible for l.h.s. of Eqn. (15.12), $y^* \leq 0$, $y^* \in P_f$, so $y^*(E) \leq f(X)$ for all X.

. . .

Min-Norm Point and SFM

. . . proof of Thm. 15.7.1 cont.

- Now, y^* is feasible for the l.h.s. of Eqn. (15.12). This follows since, we have $y^* = x^* \wedge 0 \leq 0$, and since $x^* \in B_f \subset P_f$, and $y^* \leq x^*$ and P_f is down-closed, we have that $y^* \in P_f$.
- Also, for any $y \in P_f$ with $y \leq 0$ and for any $X \subseteq E$, we have $y(E) \leq y(X) \leq f(X)$.
- Hence, we have found a feasible for l.h.s. of Eqn. (15.12), $y^* \leq 0$, $y^* \in P_f$, so $y^*(E) \leq f(X)$ for all X.
- So $y^*(E) \leq \min \{f(X) | X \subseteq V\}$.

. . .

Min-Norm Point and SFM

. . . proof of Thm. 15.7.1 cont.

- Now, y^* is feasible for the l.h.s. of Eqn. (15.12). This follows since, we have $y^* = x^* \wedge 0 \leq 0$, and since $x^* \in B_f \subset P_f$, and $y^* \leq x^*$ and P_f is down-closed, we have that $y^* \in P_f$.
- Also, for any $y \in P_f$ with $y \leq 0$ and for any $X \subseteq E$, we have $y(E) \leq y(X) \leq f(X)$.
- Hence, we have found a feasible for l.h.s. of Eqn. (15.12), $y^* \leq 0$, $y^* \in P_f$, so $y^*(E) \leq f(X)$ for all X.
- So $y^*(E) \leq \min \{f(X) | X \subseteq V\}$.
- Considering Eqn. (15.18), we have found sets A_- and A_0 with tightness in Eqn. (15.12), meaning $y^*(E) = f(A_-) = f(A_0)$.

. . .

Min-Norm Point and SFM

. . . proof of Thm. 15.7.1 cont.

- Now, y^* is feasible for the l.h.s. of Eqn. (15.12). This follows since, we have $y^* = x^* \wedge 0 \leq 0$, and since $x^* \in B_f \subset P_f$, and $y^* \leq x^*$ and P_f is down-closed, we have that $y^* \in P_f$.
- Also, for any $y \in P_f$ with $y \leq 0$ and for any $X \subseteq E$, we have $y(E) \leq y(X) \leq f(X)$.
- Hence, we have found a feasible for l.h.s. of Eqn. (15.12), $y^* \leq 0$, $y^* \in P_f$, so $y^*(E) \leq f(X)$ for all X.
- So $y^*(E) \leq \min \{f(X) | X \subseteq V\}$.
- Considering Eqn. (15.18), we have found sets A_- and A_0 with tightness in Eqn. (15.12), meaning $y^*(E) = f(A_-) = f(A_0)$.
- Hence, y^* is a maximizer of l.h.s. of Eqn. (15.12), and A_- and A_0 are minimizers of f.

. . .

Min-Norm Point and SFM

. . . proof of Thm. 15.7.1 cont.

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

$$f(X) \geq x^*(X) > x^*(A_-) = f(A_-) \tag{15.22}$$

□

Min-Norm Point and SFM

... proof of Thm. 15.7.1 cont.

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

$$f(X) \geq x^*(X) > x^*(A_-) = f(A_-) \tag{15.22}$$

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

$$f(X) \geq x^*(X) > x^*(A_0) = f(A_0) \tag{15.23}$$

□

Min-Norm Point and SFM

. . . proof of Thm. 15.7.1 cont.

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

$$f(X) \geq x^*(X) > x^*(A_-) = f(A_-) \tag{15.22}$$

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

$$f(X) \geq x^*(X) > x^*(A_0) = f(A_0) \tag{15.23}$$

- Hence, A_- must be the unique minimal minimizer of f, and A_0 is the unique maximal minimizer of f.

$\square$

Min-Norm Point and SFM

- So, if we have a procedure to compute the min-norm point computation, we can solve SFM.

Min-Norm Point and SFM

- So, if we have a procedure to compute the min-norm point computation, we can solve SFM.
- Nice thing about previous proof is that it uses both expressions for dep for different purposes.

Min-Norm Point and SFM

- So, if we have a procedure to compute the min-norm point computation, we can solve SFM.
- Nice thing about previous proof is that it uses both expressions for dep for different purposes.
- This was discovered by Fujishige (in fact the proof above is an expanded version of the one found in the book).

Min-Norm Point and SFM

- So, if we have a procedure to compute the min-norm point computation, we can solve SFM.
- Nice thing about previous proof is that it uses both expressions for dep for different purposes.
- This was discovered by Fujishige (in fact the proof above is an expanded version of the one found in the book).
- An algorithm (by F. Wolfe) can find this min-norm point, essentially an active-set procedure for quadratic programming. It uses Edmonds's greedy algorithm to make it efficient.

Min-Norm Point and SFM

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

Min-Norm Point and SFM

- So, if we have a procedure to compute the min-norm point computation, we can solve SFM.
- Nice thing about previous proof is that it uses both expressions for dep for different purposes.
- This was discovered by Fujishige (in fact the proof above is an expanded version of the one found in the book).
- An algorithm (by F. Wolfe) can find this min-norm point, essentially an active-set procedure for quadratic programming. It uses Edmonds's greedy algorithm to make it efficient.
- This is currently the best practical algorithm for general purpose submodular function minimization.
- But its underlying lower-bound complexity is unknown, although in practice its estimated empirical complexity runs anywhere from $O(n^3)$ to $O(n^{4.5})$ or so (see Jegelka, Lin, Bilmes (NIPS 2011)).

Min-norm point and other minimizers of f

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

Min-norm point and other minimizers of f

- Recall, that the set of minimizers of f forms a lattice.
- In fact, with x^* the min-norm point, and A_- and A_0 as defined above, we have the following theorem:

Min-norm point and other minimizers of f

- Recall, that the set of minimizers of f forms a lattice.
- In fact, with x^* the min-norm point, and A_- and A_0 as defined above, we have the following theorem:

Theorem 15.7.2

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

$$A = A_- \cup \bigcup_{a \in A_m} \text{dep}(x^*, a) \tag{15.24}$$

for some set $A_m \subseteq A_0 \setminus A_-$.

Min-norm point and other minimizers of f

proof of Thm. 15.7.2.

- If A is a minimizer, then $A_- \subseteq A \subseteq A_0$, and $f(A) = y^*(E)$ is the minimum valuation of f.

$\square$

Min-norm point and other minimizers of f

proof of Thm. 15.7.2.

- If A is a minimizer, then $A_- \subseteq A \subseteq A_0$, and $f(A) = y^*(E)$ is the minimum valuation of f.
- But $x^* \in P_f$, so $x^*(A) \leq f(A)$ and $f(A) = x^*(A_-) \leq x^*(A)$ (or alternatively, just note that $x^*(A_0 \setminus A) = 0$).

□

Min-norm point and other minimizers of f

proof of Thm. 15.7.2.

- If A is a minimizer, then $A_- \subseteq A \subseteq A_0$, and $f(A) = y^*(E)$ is the minimum valuation of f.
- But $x^* \in P_f$, so $x^*(A) \leq f(A)$ and $f(A) = x^*(A_-) \leq x^*(A)$ (or alternatively, just note that $x^*(A_0 \setminus A) = 0$).
- Hence, $x^*(A) = x^*(A_-) = f(A)$ so that A is also a tight set for x^*.

$\square$

Min-norm point and other minimizers of f

proof of Thm. 15.7.2.

- If A is a minimizer, then $A_- \subseteq A \subseteq A_0$, and $f(A) = y^*(E)$ is the minimum valuation of f.
- But $x^* \in P_f$, so $x^*(A) \leq f(A)$ and $f(A) = x^*(A_-) \leq x^*(A)$ (or alternatively, just note that $x^*(A_0 \setminus A) = 0$).
- Hence, $x^*(A) = x^*(A_-) = f(A)$ so that A is also a tight set for x^*.
- For any $a \in A$, A is a tight set containing a, and $\mathrm{dep}(x^*, a)$ is the minimal tight containing a.

□

Min-norm point and other minimizers of f

proof of Thm. 15.7.2.

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

□

Min-norm point and other minimizers of f

proof of Thm. 15.7.2.

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

$\square$

Min-norm point and other minimizers of f

proof of Thm. 15.7.2.

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

$\square$

On a unique minimizer f

- Note that if $f(e|A) > 0, \forall A \subseteq E$ and $e \in E \setminus A$, then we have $A_- = A_0$ (there is one unique minimizer).

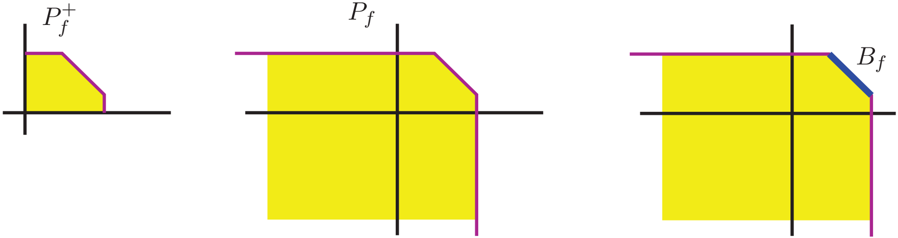
On a unique minimizer f

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

On a unique minimizer f

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

Multiple Polytopes associated with f



$$P_f^+ = P_f \cap \{x \in \mathbb{R}^E : x \geq 0\} \tag{15.5}$$

$$P_f = \{x \in \mathbb{R}^E : x(S) \leq f(S), \forall S \subseteq E\} \tag{15.6}$$

$$B_f = P_f \cap \{x \in \mathbb{R}^E : x(E) = f(E)\} \tag{15.7}$$

Polymatroidal polyhedron and greedy

- Thus, restating the above results into a single complete theorem, we have a result very similar to what we saw for matroids (i.e., Theorem **??**)

Theorem 15.8.1

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

Optimization over P_f

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

$$\begin{aligned} \text{maximize} \quad & w^\intercal x && (15.25a) \\ \text{subject to} \quad & x \in P_f && (15.25b) \end{aligned}$$

Optimization over P_f

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

$$\begin{aligned} &\text{maximize} && w^\intercal x && (15.25a) \\ &\text{subject to} && x \in P_f && (15.25b) \end{aligned}$$

- Since P_f is down closed, if $\exists e \in E$ with $w(e) < 0$ then the solution above is unboundedly large.

Optimization over P_f

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

$$\text{maximize} \quad w^\intercal x \tag{15.25a}$$

$$\text{subject to} \quad x \in P_f \tag{15.25b}$$

- Since P_f is down closed, if $\exists e \in E$ with $w(e) < 0$ then the solution above is unboundedly large. Hence, assume $w \in \mathbb{R}_+^E$.

Optimization over P_f

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

$$\begin{aligned} &\text{maximize} && w^\intercal x && (15.25a) \\ &\text{subject to} && x \in P_f && (15.25b) \end{aligned}$$

- Since P_f is down closed, if $\exists e \in E$ with $w(e) < 0$ then the solution above is unboundedly large. Hence, assume $w \in \mathbb{R}_+^E$.
- The greedy algorithm will solve this, and the proof almost identical.

Optimization over P_f

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

$$\begin{aligned} &\text{maximize} && w^\intercal x && (15.25a) \\ &\text{subject to} && x \in P_f && (15.25b) \end{aligned}$$

- Since P_f is down closed, if $\exists e \in E$ with $w(e) < 0$ then the solution above is unboundedly large. Hence, assume $w \in \mathbb{R}_+^E$.
- The greedy algorithm will solve this, and the proof almost identical.
- Due to Theorem **??**, any $x \in P_f$ with $x \notin B_f$ is dominated by $x \leq y \in B_f$ which can only increase $w^\intercal x \leq w^\intercal y$.

Optimization over P_f

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

$$\begin{aligned} &\text{maximize} && w^\intercal x && (15.25a)\\ &\text{subject to} && x \in P_f && (15.25b)\end{aligned}$$

- Since P_f is down closed, if $\exists e \in E$ with $w(e) < 0$ then the solution above is unboundedly large. Hence, assume $w \in \mathbb{R}_+^E$.
- The greedy algorithm will solve this, and the proof almost identical.
- Due to Theorem **??**, any $x \in P_f$ with $x \notin B_f$ is dominated by $x \leq y \in B_f$ which can only increase $w^\intercal x \leq w^\intercal y$.
- Hence, the problem is equivalent to: given $w \in \mathbb{R}_+^E$,

$$\begin{aligned} &\text{maximize} && w^\intercal x && (15.26a)\\ &\text{subject to} && x \in B_f && (15.26b)\end{aligned}$$

Optimization over P_f

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

$$\begin{aligned} &\text{maximize} && w^\intercal x && (15.25a)\\ &\text{subject to} && x \in P_f && (15.25b) \end{aligned}$$

- Since P_f is down closed, if $\exists e \in E$ with $w(e) < 0$ then the solution above is unboundedly large. Hence, assume $w \in \mathbb{R}_+^E$.
- The greedy algorithm will solve this, and the proof almost identical.
- Due to Theorem **??**, any $x \in P_f$ with $x \notin B_f$ is dominated by $x \leq y \in B_f$ which can only increase $w^\intercal x \leq w^\intercal y$.
- Hence, the problem is equivalent to: given $w \in \mathbb{R}_+^E$,

$$\begin{aligned} &\text{maximize} && w^\intercal x && (15.26a)\\ &\text{subject to} && x \in B_f && (15.26b) \end{aligned}$$

- 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 $w \in \mathbb{R}^E$,

$$\text{maximize} \quad w^\intercal x \tag{15.27a}$$
$$\text{subject to} \quad x \in P_f \tag{15.27b}$$

A continuous extension of f

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

$$\begin{aligned} &\text{maximize} && w^\intercal x && (15.27a)\\ &\text{subject to} && x \in P_f && (15.27b) \end{aligned}$$

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

$$\tilde{f}(w) = \max(wx : x \in P_f) \tag{15.28}$$

A continuous extension of f

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

$$\begin{aligned} &\text{maximize} && w^\intercal x && (15.27a) \\ &\text{subject to} && x \in P_f && (15.27b) \end{aligned}$$

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

$$\tilde{f}(w) = \max(wx : x \in P_f) \tag{15.28}$$

- Hence, for any w, from the above theorem, we can compute the value of this function using the greedy algorithm (after of course checking for $w \in \mathbb{R}_+^E$).

A continuous extension of f

- That is, given a submodular function f, a $w \in \mathbb{R}^E$, and defining $E_i = \{e_1, e_2, \dots, e_i\}$ and where we choose the element order $(e_1, e_2, \dots, e_m)$ based on decreasing w,so that $w(e_1) \geq w(e_2) \geq \dots \geq w(e_m)$, we have

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

A continuous extension of f

- That is, given a submodular function f, a $w \in \mathbb{R}^E$, and defining $E_i = \{e_1, e_2, \dots, e_i\}$ and where we choose the element order $(e_1, e_2, \dots, e_m)$ based on decreasing w,so that $w(e_1) \geq w(e_2) \geq \dots \geq w(e_m)$, we have

$$\tilde{f}(w) = \max(wx : x \in P_f) \tag{15.29}$$

A continuous extension of f

- That is, given a submodular function f, a $w \in \mathbb{R}^E$, and defining $E_i = \{e_1, e_2, \dots, e_i\}$ and where we choose the element order $(e_1, e_2, \dots, e_m)$ based on decreasing w,so that $w(e_1) \geq w(e_2) \geq \cdots \geq w(e_m)$, we have

$$\tilde{f}(w) = \max(wx : x \in P_f) \tag{15.29}$$

$$= \sum_{i=1}^{m} w(e_i) f(e_i | E_{i-1}) \tag{15.30}$$

A continuous extension of f

- That is, given a submodular function f, a $w \in \mathbb{R}^E$, and defining $E_i = \{e_1, e_2, \dots, e_i\}$ and where we choose the element order $(e_1, e_2, \dots, e_m)$ based on decreasing w,so that $w(e_1) \geq w(e_2) \geq \dots \geq w(e_m)$, we have

$$\tilde{f}(w) = \max(wx : x \in P_f) \tag{15.29}$$

$$= \sum_{i=1}^{m} w(e_i) f(e_i | E_{i-1}) \tag{15.30}$$

$$= \sum_{i=1}^{m} w(e_i)(f(E_i) - f(E_{i-1})) \tag{15.31}$$

A continuous extension of f

- That is, given a submodular function f, a $w \in \mathbb{R}^E$, and defining $E_i = \{e_1, e_2, \dots, e_i\}$ and where we choose the element order $(e_1, e_2, \dots, e_m)$ based on decreasing w,so that $w(e_1) \geq w(e_2) \geq \dots \geq w(e_m)$, we have

$$\tilde{f}(w) = \max(wx : x \in P_f) \tag{15.29}$$

$$= \sum_{i=1}^{m} w(e_i) f(e_i | E_{i-1}) \tag{15.30}$$

$$= \sum_{i=1}^{m} w(e_i)(f(E_i) - f(E_{i-1})) \tag{15.31}$$

$$= w(e_m) f(E_m) + \sum_{i=1}^{m-1} (w(e_i) - w(e_{i+1})) f(E_i) \tag{15.32}$$

A continuous extension of f

- That is, given a submodular function f, a $w \in \mathbb{R}^E$, and defining $E_i = \{e_1, e_2, \dots, e_i\}$ and where we choose the element order $(e_1, e_2, \dots, e_m)$ based on decreasing w,so that $w(e_1) \geq w(e_2) \geq \dots \geq w(e_m)$, we have

$$\tilde{f}(w) = \max(wx : x \in P_f) \tag{15.29}$$
$$= \sum_{i=1}^{m} w(e_i) f(e_i | E_{i-1}) \tag{15.30}$$
$$= \sum_{i=1}^{m} w(e_i)(f(E_i) - f(E_{i-1})) \tag{15.31}$$
$$= w(e_m) f(E_m) + \sum_{i=1}^{m-1} (w(e_i) - w(e_{i+1})) f(E_i) \tag{15.32}$$

- We say that $\emptyset \triangleq E_0 \subset E_1 \subset E_2 \subset \dots \subset E_m = E$ forms a chain based on w.

A continuous extension of f

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

$$\tilde{f}(w) = \max(wx : x \in P_f) \tag{15.33}$$

A continuous extension of f

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

$$\tilde{f}(w) = \max(wx : x \in P_f) \tag{15.33}$$

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

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

A continuous extension of f

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

$$\tilde{f}(w) = \max(wx : x \in P_f) \tag{15.33}$$

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

$$\tilde{f}(w) = w(e_m)f(E_m) + \sum_{i=1}^{m-1} (w(e_i) - w(e_{i+1}))f(E_i) \tag{15.34}$$

A continuous extension of f

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

$$\tilde{f}(w) = \max(wx : x \in P_f) \tag{15.33}$$

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

$$\tilde{f}(w) = w(e_m)f(E_m) + \sum_{i=1}^{m-1}(w(e_i) - w(e_{i+1}))f(E_i) \tag{15.34}$$

$$= \sum_{i=1}^{m} \lambda_i f(E_i) \tag{15.35}$$

A continuous extension of f

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

$$\tilde{f}(w) = \max(wx : x \in P_f) \tag{15.33}$$

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

$$\tilde{f}(w) = w(e_m)f(E_m) + \sum_{i=1}^{m-1}(w(e_i) - w(e_{i+1}))f(E_i) \tag{15.34}$$

$$= \sum_{i=1}^{m} \lambda_i f(E_i) \tag{15.35}$$

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

A continuous extension of f

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

$$\tilde{f}(w) = \max(wx : x \in P_f) \tag{15.33}$$

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

$$\tilde{f}(w) = w(e_m)f(E_m) + \sum_{i=1}^{m-1}(w(e_i) - w(e_{i+1}))f(E_i) \tag{15.34}$$

$$= \sum_{i=1}^{m} \lambda_i f(E_i) \tag{15.35}$$

  where $\lambda_m = w(e_m)$ and otherwise $\lambda_i = w(e_i) - w(e_{i+1})$, where the elements are sorted according to w as before.
- From convex analysis, we know $\tilde{f}(w) = \max(wx : x \in P)$ is always convex in w for any set $P \subseteq R^E$, since it is the maximum of a set of linear functions (true even when f is not submodular or P is not a convex set).

An extension of f

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

$$\begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} = \underbrace{(w_1 - w_2)}_{\lambda_1} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \underbrace{(w_2 - w_3)}_{\lambda_2} \begin{pmatrix} 1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \cdots + \underbrace{(w_{n-1} - w_n)}_{\lambda_{m-1}} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 0 \end{pmatrix} + \underbrace{(w_m)}_{\lambda_m} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{pmatrix} \tag{15.36}$$

An extension of f

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

$$\begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} = \underbrace{(w_1 - w_2)}_{\lambda_1} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \underbrace{(w_2 - w_3)}_{\lambda_2} \begin{pmatrix} 1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \cdots + \underbrace{(w_{n-1} - w_n)}_{\lambda_{m-1}} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 0 \end{pmatrix} + \underbrace{(w_m)}_{\lambda_m} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{pmatrix} \tag{15.36}$$

- If we take w in decreasing order, then each coefficient of the vectors is non-negative (except possibly the last one, $\lambda_m = w_m$).

An extension of f

- Define sets E_i based on this decreasing order of w as follows, for $i = 0, \dots, n$

$$E_i \stackrel{\text{def}}{=} \{e_1, e_2, \dots, e_i\} \tag{15.37}$$

An extension of f

- Define sets E_i based on this decreasing order of w as follows, for $i = 0, \ldots, n$

$$E_i \stackrel{\text{def}}{=} \{e_1, e_2, \ldots, e_i\} \tag{15.37}$$

- Note that

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

An extension of f

- Define sets E_i based on this decreasing order of w as follows, for $i = 0, \ldots, n$

$$E_i \stackrel{\text{def}}{=} \{e_1, e_2, \ldots, e_i\} \tag{15.37}$$

- Note 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 $\tilde{f}$ back to f, even when f is not submodular

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

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

- From the continuous $\tilde{f}$, 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 $\tilde{f}$ back to f, even when f is not submodular

- From the continuous $\tilde{f}$, 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 the elements of E in decreasing order of w so that $w(e_1) \geq w(e_2) \geq w(e_3) \geq \cdots \geq w(e_m)$.

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

- From the continuous $\tilde{f}$, 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 the elements of E in decreasing order of w so that $w(e_1) \geq w(e_2) \geq w(e_3) \geq \dots \geq w(e_m)$.
- This means

$$w = (w(e_1), w(e_2), \dots, w(e_m)) = (\underbrace{1, 1, 1, \dots, 1}_{|A| \text{ times}}, \underbrace{0, 0, \dots, 0}_{m-|A| \text{ times}}) \quad (15.38)$$

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

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

- From the continuous $\tilde{f}$, 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 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

$$w = (w(e_1), w(e_2), \ldots, w(e_m)) = (\underbrace{1, 1, 1, \ldots, 1}_{|A| \text{ times}}, \underbrace{0, 0, \ldots, 0}_{m-|A| \text{ times}}) \tag{15.38}$$

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 $E_{|A|} = \{e_1, e_2, \ldots, e_{|A|}\} = A$:

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

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

- From the continuous $\tilde{f}$, 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 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

$$w = (w(e_1), w(e_2), \ldots, w(e_m)) = (\underbrace{1, 1, 1, \ldots, 1}_{|A| \text{ times}}, \underbrace{0, 0, \ldots, 0}_{m-|A| \text{ times}}) \quad (15.38)$$

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 $E_{|A|} = \{e_1, e_2, \ldots, e_{|A|}\} = A$:

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

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

- From the continuous $\tilde{f}$, 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 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

$$w = (w(e_1), w(e_2), \dots, w(e_m)) = (\underbrace{1, 1, 1, \dots, 1}_{|A| \text{ times}}, \underbrace{0, 0, \dots, 0}_{m-|A| \text{ times}}) \tag{15.38}$$

  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 $E_{|A|} = \{e_1, e_2, \dots, e_{|A|}\} = A$:

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

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

- From the continuous $\tilde{f}$, 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 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

$$w = (w(e_1), w(e_2), \ldots, w(e_m)) = (\underbrace{1, 1, 1, \ldots, 1}_{|A| \text{ times}}, \underbrace{0, 0, \ldots, 0}_{m-|A| \text{ times}}) \tag{15.38}$$

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 $E_{|A|} = \{e_1, e_2, \ldots, e_{|A|}\} = A$:

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

$$= \mathbf{1}_A(m) f(E_m) + \sum_{i=1}^{m-1} (\mathbf{1}_A(i) - \mathbf{1}_A(i+1)) f(E_i) \tag{15.39}$$

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

- From the continuous $\tilde{f}$, 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 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

$$w = (w(e_1), w(e_2), \ldots, w(e_m)) = (\underbrace{1, 1, 1, \ldots, 1}_{|A| \text{ times}}, \underbrace{0, 0, \ldots, 0}_{m-|A| \text{ times}}) \tag{15.38}$$

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 $E_{|A|} = \{e_1, e_2, \ldots, e_{|A|}\} = A$:

$$\begin{aligned}\tilde{f}(w) &= \sum_{i=1}^{m} \lambda_i f(E_i) = w(e_m) f(E_m) + \sum_{i=1}^{m-1} (w(e_i) - w(e_{i+1}) f(E_i) \\ &= \mathbf{1}_A(m) f(E_m) + \sum_{i=1}^{m-1} (\mathbf{1}_A(i) - \mathbf{1}_A(i+1)) f(E_i) \qquad (15.39) \\ &= (\mathbf{1}_A(|A|) - \mathbf{1}_A(|A|+1)) f(E_{|A|}) = f(E_{|A|})\end{aligned}$$

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

- From the continuous $\tilde{f}$, 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 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

$$w = (w(e_1), w(e_2), \ldots, w(e_m)) = (\underbrace{1, 1, 1, \ldots, 1}_{|A| \text{ times}}, \underbrace{0, 0, \ldots, 0}_{m-|A| \text{ times}}) \quad (15.38)$$

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 $E_{|A|} = \{e_1, e_2, \ldots, e_{|A|}\} = A$:

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

$$= \mathbf{1}_A(m) f(E_m) + \sum_{i=1}^{m-1} (\mathbf{1}_A(i) - \mathbf{1}_A(i+1)) f(E_i) \quad (15.39)$$

$$= (\mathbf{1}_A(|A|) - \mathbf{1}_A(|A|+1)) f(E_{|A|}) = f(E_{|A|}) = f(A) \quad (15.40)$$

From $\tilde{f}$ back to f

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

From $\tilde{f}$ back to f

- We can view $\tilde{f} : [0,1]^E \to \mathbb{R}$ defined on the hypercube, with f defined as $\tilde{f}$ evaluated on the hypercube extreme points (vertices).
- To summarize, with $\tilde{f}(A) = \sum_{i=1}^{m} \lambda_i f(E_i)$, we have

$$\tilde{f}(\mathbf{1}_A) = f(A), \tag{15.41}$$

From $\tilde{f}$ back to f

- We can view $\tilde{f} : [0,1]^E \to \mathbb{R}$ defined on the hypercube, with f defined as $\tilde{f}$ evaluated on the hypercube extreme points (vertices).
- To summarize, with $\tilde{f}(A) = \sum_{i=1}^m \lambda_i f(E_i)$, we have

$$\tilde{f}(\mathbf{1}_A) = f(A), \tag{15.41}$$

- . . . and when f is submodular, we also have have

$$\tilde{f}(\mathbf{1}_A) = \max\{\mathbf{1}_A x : x \in P_f\}. \tag{15.42}$$

An extension of f

- Thus, for any $f : 2^E \to \mathbb{R}$, even non-submodular f, we can define an extension in this way, with

$$\tilde{f}(w) = \sum_{i=1}^{m} \lambda_i f(E_i) \tag{15.43}$$

with the $E_i = \{e_1, \ldots, e_i\}$'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

$$\text{for } i \in \{1, \ldots, m\}, \quad \lambda_i = \begin{cases} w(e_i) - w(e_{i+1}) & \text{if } i < m \\ w(e_m) & \text{if } i = m \end{cases} \tag{15.44}$$

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

An extension of f

- Thus, for any $f : 2^E \to \mathbb{R}$, even non-submodular f, we can define an extension in this way, with

$$\tilde{f}(w) = \sum_{i=1}^{m} \lambda_i f(E_i) \tag{15.43}$$

with the $E_i = \{e_1, \ldots, e_i\}$'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

$$\text{for } i \in \{1, \ldots, m\}, \quad \lambda_i = \begin{cases} w(e_i) - w(e_{i+1}) & \text{if } i < m \\ w(e_m) & \text{if } i = m \end{cases} \tag{15.44}$$

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

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

Weighted gains vs. weighted functions

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

$$\tilde{f}(w) = \sum_{i=1}^{m} w(e_i) f(e_i | E_{i-1}) \quad (15.45)$$

$$= \sum_{i=1}^{m} w(e_i)(f(E_i) - f(E_{i-1})) \quad (15.46)$$

$$= w(e_m) f(E_m) + \sum_{i=1}^{m-1} (w(e_i) - w(e_{i+1})) f(E_i) \quad (15.47)$$

$$= \sum_{i=1}^{m} \lambda_i f(E_i) \quad (15.48)$$

Weighted gains vs. weighted functions

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

$$\begin{aligned}
\tilde{f}(w) &= \sum_{i=1}^{m} w(e_i) f(e_i | E_{i-1}) && (15.45) \\
&= \sum_{i=1}^{m} w(e_i)(f(E_i) - f(E_{i-1})) && (15.46) \\
&= w(e_m) f(E_m) + \sum_{i=1}^{m-1} (w(e_i) - w(e_{i+1})) f(E_i) && (15.47) \\
&= \sum_{i=1}^{m} \lambda_i f(E_i) && (15.48)
\end{aligned}$$

- So $\tilde{f}(w)$ seen either as sum of weighted gain evaluatiosn (Eqn. (15.45), or as sum of weighted function evaluations (Eqn. (15.48)).

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

- Lovász showed that if a function $\tilde{f}(w)$ defined as in Eqn. (15.43) is convex, then f must be submodular.

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

- Lovász showed that if a function $\tilde{f}(w)$ defined as in Eqn. (15.43) is convex, then f must be submodular.
- This continuous extension $\tilde{f}$ of f, in any case (f being submodular or not), is called the Lovász extension of f.

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

- Lovász showed that if a function $\tilde{f}(w)$ defined as in Eqn. (15.43) is convex, then f must be submodular.
- This continuous extension $\tilde{f}$ of f, in any case (f being submodular or not), is called the Lovász extension of f.
- Note, also possible to define this when $f(\emptyset) \neq 0$ (but doesn't really add any generality).

Lovász Extension, Submodularity and Convexity

Theorem 15.8.1

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

Proof.

- We've already seen that if f is submodular, its extension can be written via Eqn.(15.43) due to the greedy algorithm, and therefore is also equivalent to $\tilde{f}(w) = \max\{wx : x \in P_f\}$, and thus is convex.

...

Lovász Extension, Submodularity and Convexity

Theorem 15.8.1

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

Proof.

- We've already seen that if f is submodular, its extension can be written via Eqn.(15.43) due to the greedy algorithm, and therefore is also equivalent to $\tilde{f}(w) = \max\{wx : x \in P_f\}$, and thus is convex.
- Conversely, suppose the Lovász extension $\tilde{f}(w) = \sum_i \lambda_i f(E_i)$ of some function $f : 2^E \to \mathbb{R}$ is a convex function.

...

Lovász Extension, Submodularity and Convexity

Theorem 15.8.1

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

Proof.

- We've already seen that if f is submodular, its extension can be written via Eqn.(15.43) due to the greedy algorithm, and therefore is also equivalent to $\tilde{f}(w) = \max \{wx : x \in P_f\}$, and thus is convex.
- Conversely, suppose the Lovász extension $\tilde{f}(w) = \sum_i \lambda_i f(E_i)$ 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 $\tilde{f}(\alpha w) = \alpha \tilde{f}(w)$ for any $\alpha \in \mathbb{R}_+$. I.e., f is a positively homogeneous convex function.

...

Lovász Extension, Submodularity and Convexity

. . . proof of Thm. 15.8.1 cont.

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

. . .

Lovász Extension, Submodularity and Convexity

. . . proof of Thm. 15.8.1 cont.

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

$$\tilde{f}(\mathbf{1}_A + \mathbf{1}_B) = \tilde{f}(\mathbf{1}_{A \cup B} + \mathbf{1}_{A \cap B}) \tag{15.49}$$

$$= f(A \cup B) + f(A \cap B). \tag{15.50}$$

...

Lovász Extension, Submodularity and Convexity

. . . proof of Thm. 15.8.1 cont.

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

$$\tilde{f}(\mathbf{1}_A + \mathbf{1}_B) = \tilde{f}(\mathbf{1}_{A\cup B} + \mathbf{1}_{A\cap B}) \tag{15.49}$$

$$= f(A \cup B) + f(A \cap B). \tag{15.50}$$

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

$$w = (w(e_1), w(e_2), \dots, w(e_m)) \tag{15.51}$$

$$= (\underbrace{2, 2, \dots, 2}_{i\in C}, \underbrace{1, 1, \dots, 1}_{i\in A\triangle B}, \underbrace{0, 0, \dots, 0}_{i\in E\setminus(A\cup B)}) \tag{15.52}$$

. . .

Lovász Extension, Submodularity and Convexity

. . . proof of Thm. 15.8.1 cont.

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

$$\tilde{f}(\mathbf{1}_A + \mathbf{1}_B) = \tilde{f}(\mathbf{1}_{A\cup B} + \mathbf{1}_{A\cap B}) \tag{15.49}$$

$$= f(A \cup B) + f(A \cap B). \tag{15.50}$$

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

$$w = (w(e_1), w(e_2), \dots, w(e_m)) \tag{15.51}$$

$$= (\underbrace{2, 2, \dots, 2}_{i\in C}, \underbrace{1, 1, \dots, 1}_{i\in A\triangle B}, \underbrace{0, 0, \dots, 0}_{i\in E\setminus(A\cup B)}) \tag{15.52}$$

- Then, considering $\tilde{f}(w) = \sum_i \lambda_i f(E_i)$, 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. 15.8.1 cont.

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

$$\tilde{f}(\mathbf{1}_A + \mathbf{1}_B) = \tilde{f}(\mathbf{1}_{A\cup B} + \mathbf{1}_{A\cap B}) \tag{15.49}$$

$$= f(A \cup B) + f(A \cap B). \tag{15.50}$$

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

$$w = (w(e_1), w(e_2), \dots, w(e_m)) \tag{15.51}$$

$$= (\underbrace{2, 2, \dots, 2}_{i \in C}, \underbrace{1, 1, \dots, 1}_{i \in A \triangle B}, \underbrace{0, 0, \dots, 0}_{i \in E \setminus (A \cup B)}) \tag{15.52}$$

- Then, considering $\tilde{f}(w) = \sum_i \lambda_i f(E_i)$, 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, $\tilde{f}(w) = f(A \cap B) + f(A \cup B)$.

. . .

Lovász Extension, Submodularity and Convexity

. . . proof of Thm. 15.8.1 cont.

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

$$0.5[f(A \cap B) + f(A \cup B)] \tag{15.56}$$

□

Lovász Extension, Submodularity and Convexity

. . . proof of Thm. 15.8.1 cont.

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

$$0.5[f(A \cap B) + f(A \cup B)] = 0.5[\tilde{f}(\mathbf{1}_A + \mathbf{1}_B)] \tag{15.53}$$

(15.56)

□

Lovász Extension, Submodularity and Convexity

. . . proof of Thm. 15.8.1 cont.

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

$$0.5[f(A \cap B) + f(A \cup B)] = 0.5[\tilde{f}(\mathbf{1}_A + \mathbf{1}_B)] \tag{15.53}$$
$$= \tilde{f}(0.5\mathbf{1}_A + 0.5\mathbf{1}_B) \tag{15.54}$$

$$\tag{15.56}$$

□

Lovász Extension, Submodularity and Convexity

. . . proof of Thm. 15.8.1 cont.

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

$$0.5[f(A \cap B) + f(A \cup B)] = 0.5[\tilde{f}(\mathbf{1}_A + \mathbf{1}_B)] \tag{15.53}$$

$$= \tilde{f}(0.5\mathbf{1}_A + 0.5\mathbf{1}_B) \tag{15.54}$$

$$\leq 0.5\tilde{f}(\mathbf{1}_A) + 0.5\tilde{f}(\mathbf{1}_B) \tag{15.55}$$

$$\tag{15.56}$$

□

Lovász Extension, Submodularity and Convexity

. . . proof of Thm. 15.8.1 cont.

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

$$0.5[f(A \cap B) + f(A \cup B)] = 0.5[\tilde{f}(\mathbf{1}_A + \mathbf{1}_B)] \tag{15.53}$$
$$= \tilde{f}(0.5\mathbf{1}_A + 0.5\mathbf{1}_B) \tag{15.54}$$
$$\leq 0.5\tilde{f}(\mathbf{1}_A) + 0.5\tilde{f}(\mathbf{1}_B) \tag{15.55}$$
$$= 0.5(f(A) + f(B)) \tag{15.56}$$

□

Lovász Extension, Submodularity and Convexity

. . . proof of Thm. 15.8.1 cont.

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

$$0.5[f(A \cap B) + f(A \cup B)] = 0.5[\tilde{f}(\mathbf{1}_A + \mathbf{1}_B)] \tag{15.53}$$

$$= \tilde{f}(0.5\mathbf{1}_A + 0.5\mathbf{1}_B) \tag{15.54}$$

$$\leq 0.5\tilde{f}(\mathbf{1}_A) + 0.5\tilde{f}(\mathbf{1}_B) \tag{15.55}$$

$$= 0.5(f(A) + f(B)) \tag{15.56}$$

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

$$f(A \cup B) + f(A \cap B) \leq f(A) + f(B) \tag{15.57}$$

so f must be submodular.

□

Edmonds - Submodularity - 1969

SUBMODULAR FUNCTIONS, MATROIDS, AND CERTAIN POLYHEDRA*

Jack Edmonds

National Bureau of Standards, Washington, D.C.,U.S.A.

I.

The viewpoint of the subject of matroids, and related areas of lattice theory, has always been, in one way or another, abstraction of algebraic dependence or, equivalently, abstraction of the incidence relations in geometric representations of algebra. Often one of the