Submodular Functions, Optimization, and Applications to Machine Learning

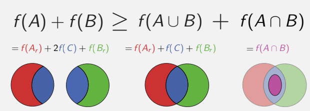
— Spring Quarter, Lecture 15 —

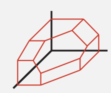
http://j.ee.washington.edu/~bilmes/classes/ee596b_spring_2014/

Prof. Jeff Bilmes

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

May 19th, 2014





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, polymatroids and entropic Venn diagrams, exchange capacity,
- L16: minimum norm point algorithm and the lattice of minimizers of a submodular function, Lovasz extension
- 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

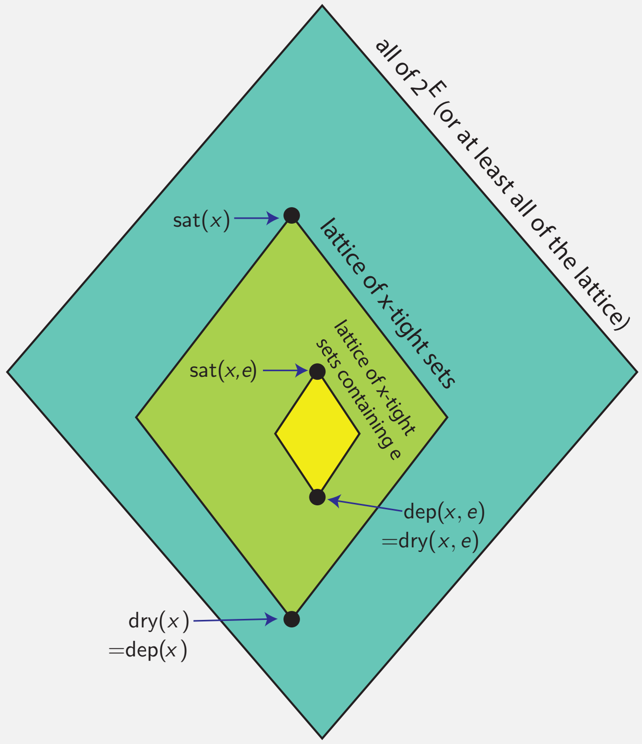
- x-tight sets: For $x \in P_f$, $\mathcal{D}(x) = \{A \subseteq E : x(A) = f(A)\}$.
- Polymatroid closure/maximal x-tight set: For $x \in P_f$, $\mathrm{sat}(x) = \cup\{A : A \in \mathcal{D}(x)\} = \{e : e \in E, \forall \alpha > 0, x + \alpha \mathbf{1}_e \notin P_f\}$.
- Saturation capacity: for $x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

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

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

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, $\bigcap_e \text{dep}(x, e) = \text{dep}(x)$.

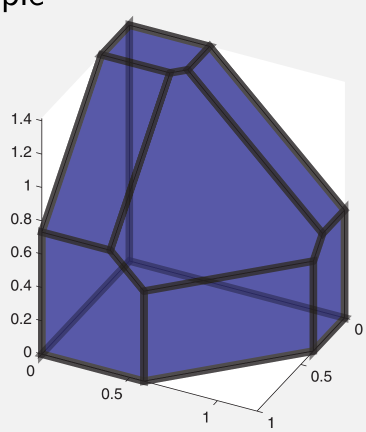


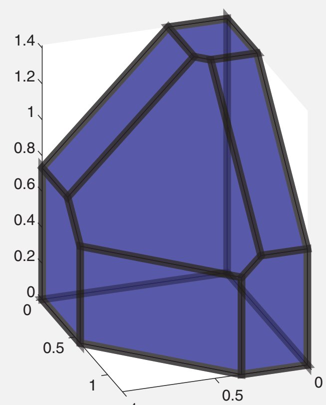
Support of vector

- 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.
- $\text{supp}(x)$ 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 $1 \le i \le k \le |E|$, 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 $1 \le i \le k$, if $e_i \notin \text{supp}(x)$, $x(E_k) = x(E_k - e_i)$
  3. Also, for $1 \le i \le k$, if $e_i \notin \text{supp}(x)$, then $x(e_i) = 0 = f(e_i|E_{i-1}) \ge f(e_i|E_k - e_i) = f(E_k|E_k - e_i) \ge 0$ 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 $\text{supp}(x) = \text{sat}(x)$?
- 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 $x(Y) = 0$.
- Suppose $\text{supp}(x) = X$, 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, $\text{sat}(x) \supseteq X \cup Y$. Hence, $\text{sat}(x) \supset \text{supp}(x)$.
- 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 sat, and dep

- 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,

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

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

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

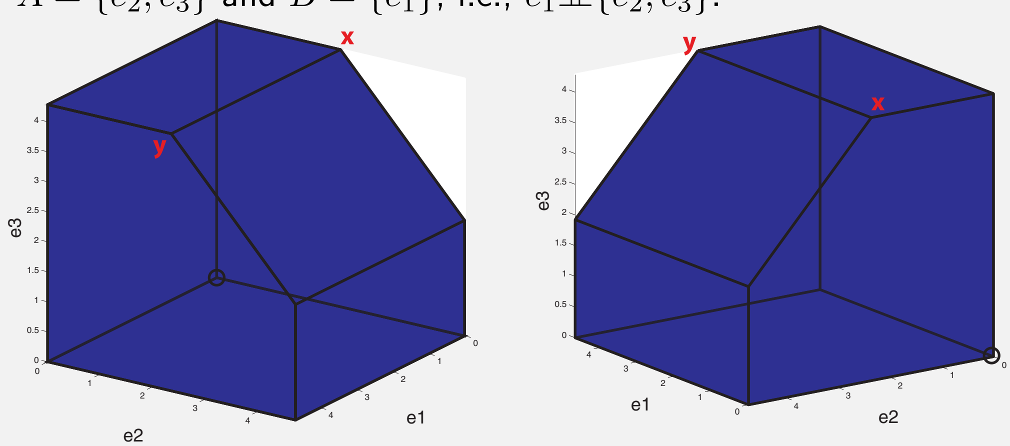
- For $e \in \text{sat}(x)$, we have $\text{dep}(x, e) \subseteq \text{sat}(x)$ (fundamental circuit) is the minimal (common) saturated (x-tight) set w.r.t. x containing e. I.e.,

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

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

supp, sat, dep, example with perfect independence

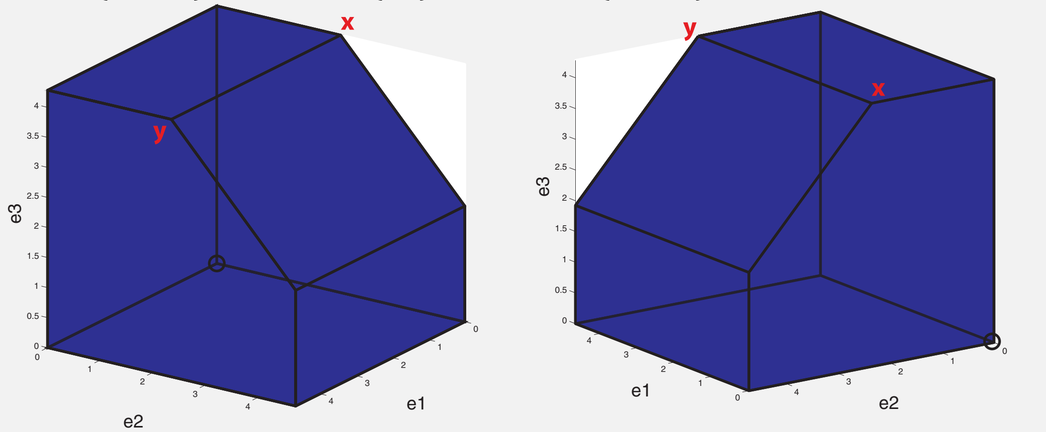
- Example polymatroid where there is perfect independence between $A = \{e_2, e_3\}$ and $B = \{e_1\}$, i.e., $e_1 \perp\!\!\!\perp \{e_2, e_3\}$.



- 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

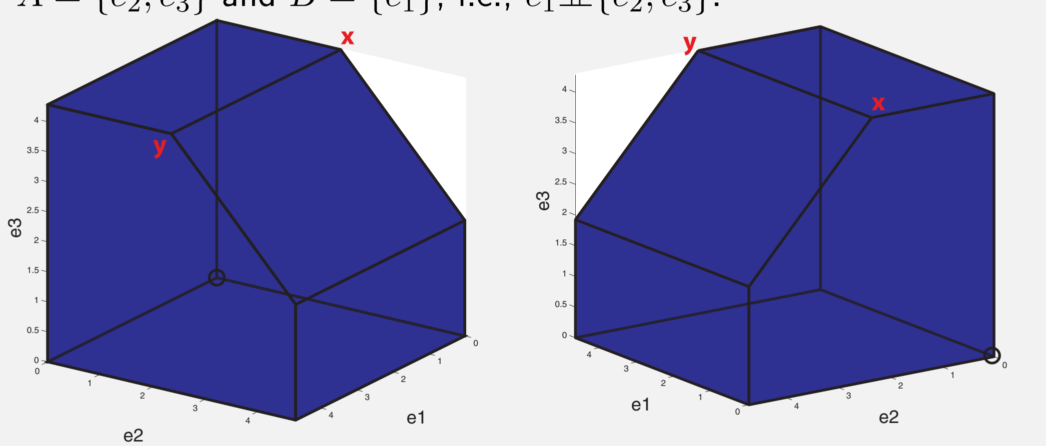
- Example polymatroid where there is perfect independence between $A = \{e_2, e_3\}$ and $B = \{e_1\}$, i.e., $e_1 \perp\!\!\!\perp \{e_2, e_3\}$.



- 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

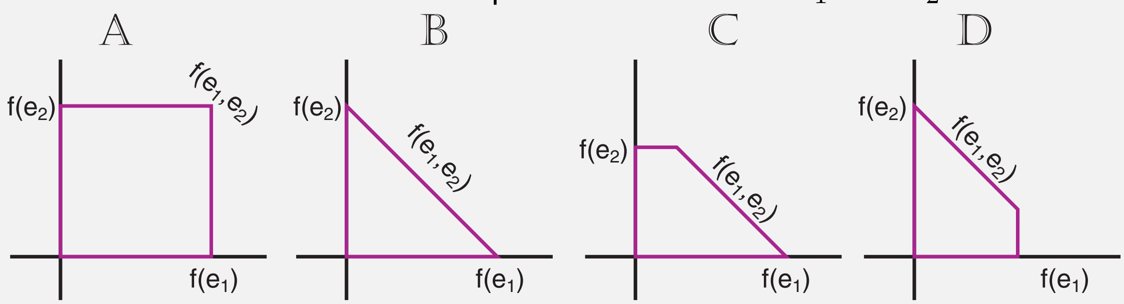
- Example polymatroid where there is perfect independence between $A = \{e_2, e_3\}$ and $B = \{e_1\}$, i.e., $e_1 \perp\!\!\!\perp \{e_2, e_3\}$.



- 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, and polymatroid dependence in 2D

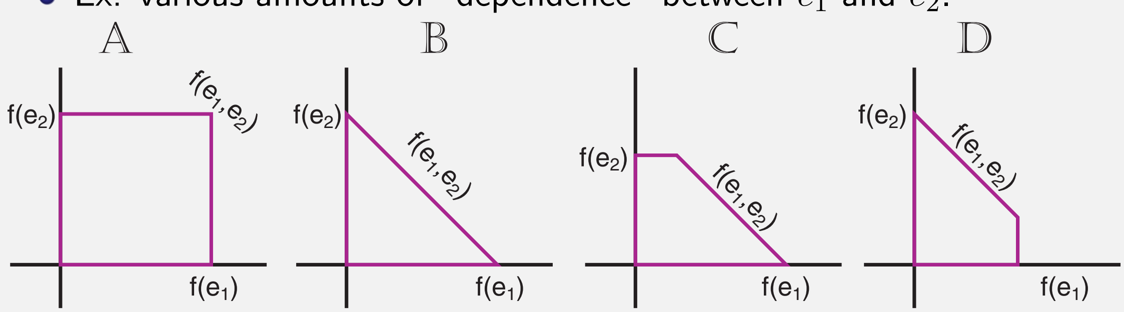
- Ex: various amounts of "dependence" between e_1 and e_2.



- 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

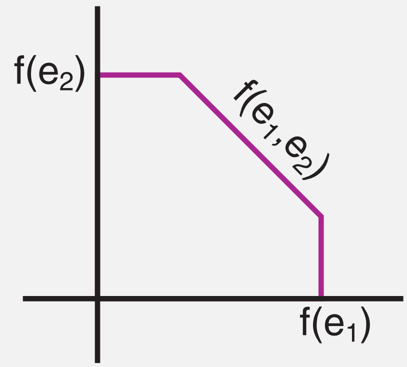
- Ex: various amounts of "dependence" between e_1 and e_2.



- 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 $x = (0, f(e_2))$.

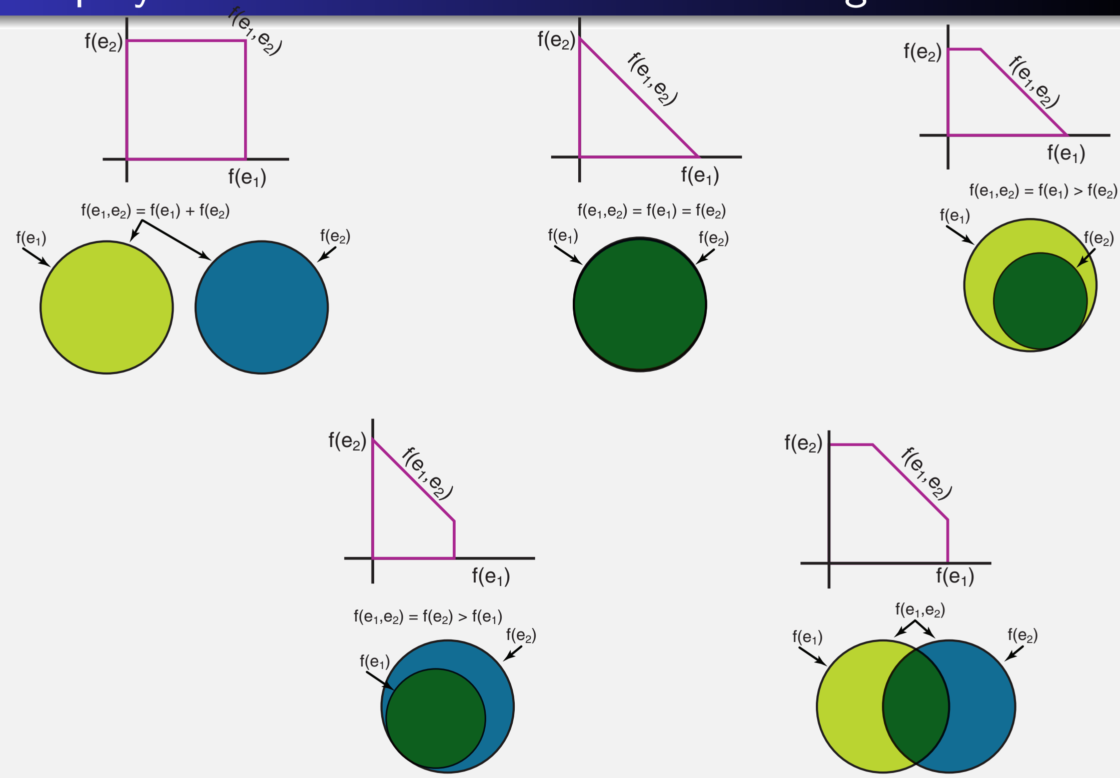
supp, sat, and dependence in 2D

- 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



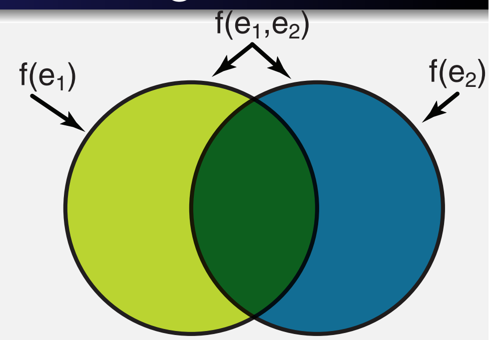
2D polymatroids and Information Venn Diagrams

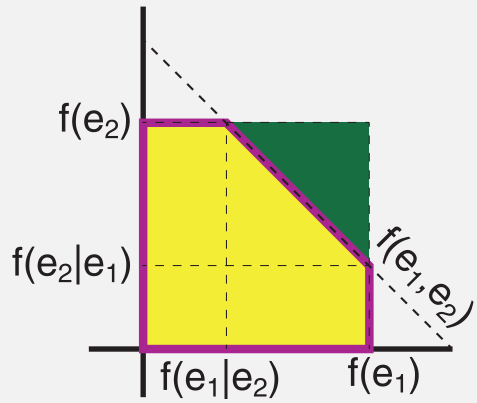
- Consider symmeterized combinatorial mutual information function:

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

- Consider area of green triangle:

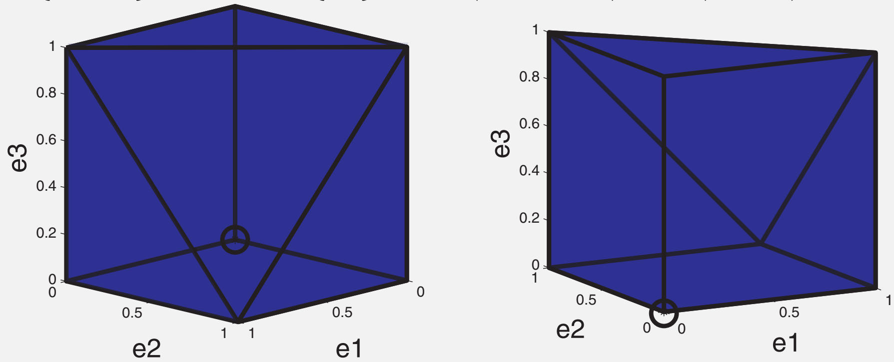
$$\frac{1}{2}\Big(f(e_2) - f(e_2|e_1)\Big)\Big(f(e_1) - f(e_1|e_2)\Big) = \frac{1}{2}(f(e_1) + f(e_2) - f(e_1, e_2))^2 = \frac{1}{2}\Big(I_f(e_1, e_2)\Big)^2$$





supp, sat, and perfect dependence in 3D

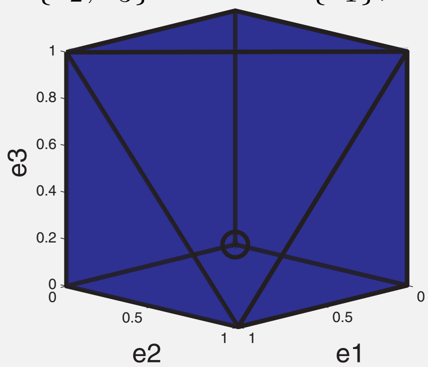
- Ex: polymatroid with perfect independence between e_2 and e_3, so $f(e_2, e_3) = f(e_2) + f(e_3)$, but perfect dependence between $A = \{e_2, e_3\}$ and $B = \{e_1\}$, so $f(e_1, e_2, e_3) = f(e_2, e_3)$

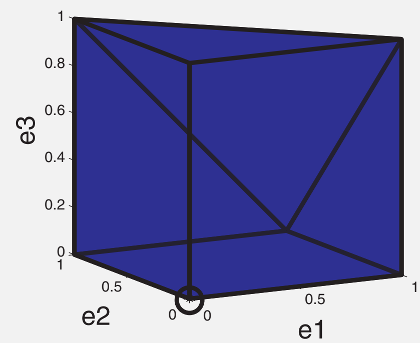


- Entropy case: xor V-structure Bayesian network $e_1 = h(e_2, e_3)$ where h is the xor function ($e_2 \to e_1 \leftarrow e_3$), and e_2, e_3 are both independent binary with unity entropy.
- Q: Why does the polytope have a symmetry? Notice independence (square) for any pair.

supp, sat, and perfect dependence in 3D

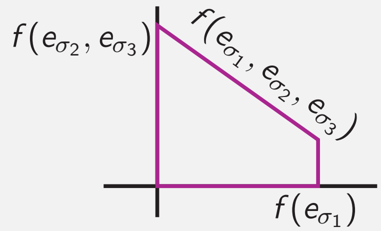
- Ex: polymatroid with perfect independence between e_2 and e_3, so $f(e_2, e_3) = f(e_2) + f(e_3)$, but perfect dependence between $A = \{e_2, e_3\}$ and $B = \{e_1\}$, so $f(e_1, e_2, e_3) = f(e_2, e_3)$





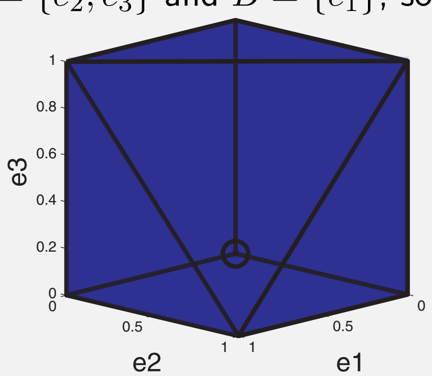
- For any permutation σ of $\{1, 2, 3\}$, considering $\{e_{\sigma_1}, e_{\sigma_2}\}$ vs. $\{e_{\sigma_3}\}$:

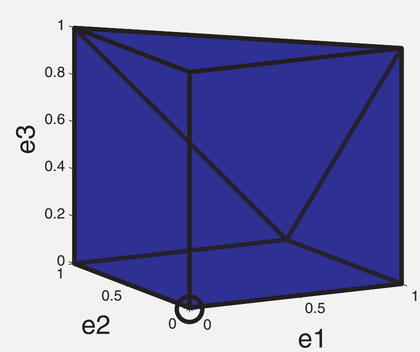
e_{σ_3} is a deterministic function of $\{e_{\sigma_1}, e_{\sigma_2}\}$



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





- 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.
- 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 $i \neq j$.
- 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

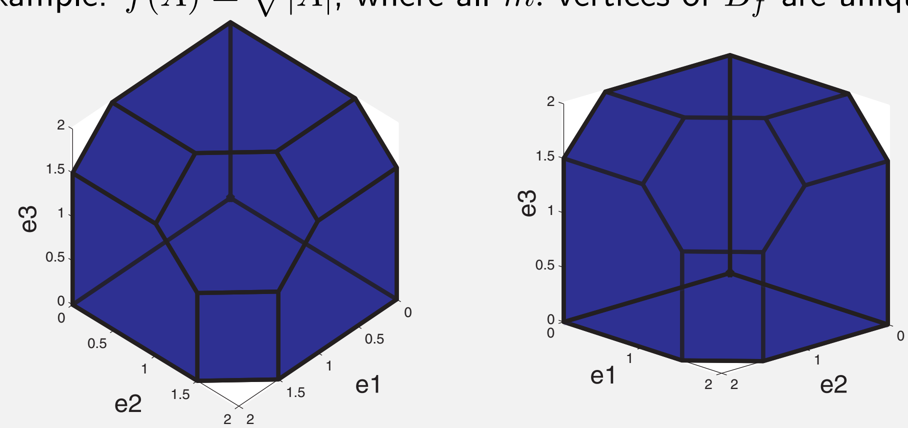
- In general, for extremal x, $\text{sat}(x) \supseteq \text{supp}(x)$.
- Now, for any $e \in E \setminus \text{supp}(x)$, we clearly have $x(\text{supp}(x) + e) = x(\text{supp}(x))$ since $x(e) = 0$.
- On the other hand, for $e_i \in \text{sat}(x) \setminus \text{supp}(x)$, we have perfect dependence, 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).
- 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 $E = (e_1, \dots)$, 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 $(E_1, E_2, \ldots, E_k)$ 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 $|E| - k$.

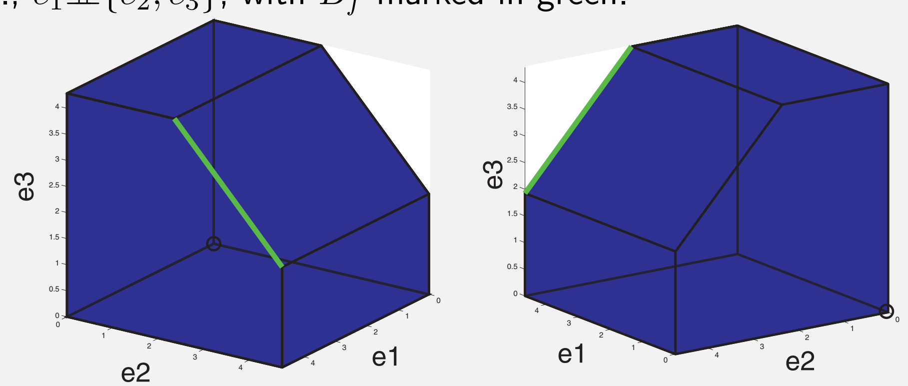
- Thus, "independence" between disjoint A and B (leading to a rectangular projection of the polymatroid polytope) reduces the dimension of the base polytope, as expected.
- Thus, any point $x \in B_f$ is a convex combination of at most $|E| - k + 1$ vertices of B_f.
- And if f does not have such independence,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

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

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

B_f dominates P_f

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

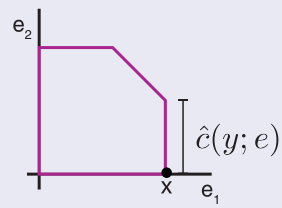
Theorem 15.5.2

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.

- 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};
```



...

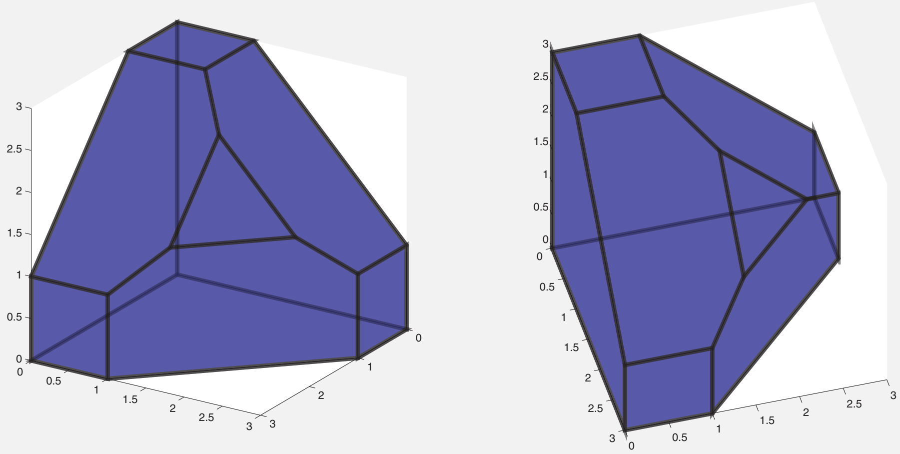
B_f dominates P_f

...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 $y(e) \leq f(T' + e) - y(T') = y(e) + f(T' + e) - y(T' + e)$.
- 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

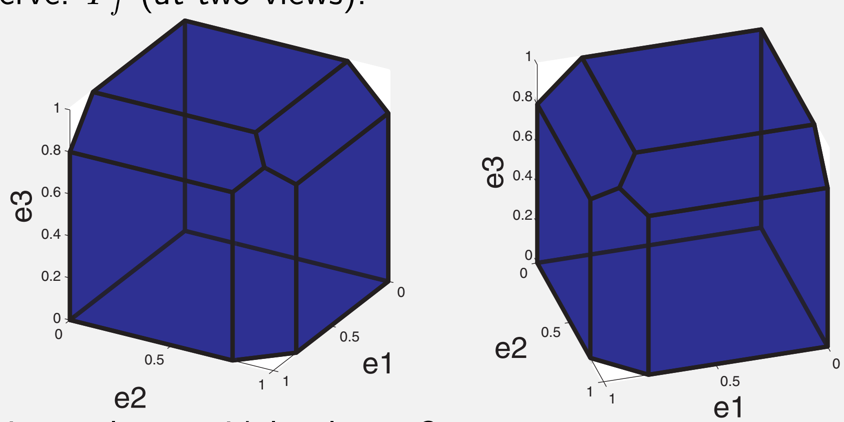
- Observe: P_f (at two views):



- 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 $f(\{e_1, e_2, e_3\}) + f(\{e_1\}) = 5.5 + 3 = 8.5$.

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.
- Thus we have for $x \in P_f$,

$$\hat{c}(x;e) \stackrel{\text{def}}{=} \min \{f(A) - x(A), \forall A \ni e\} \tag{15.23}$$
$$= \max \{\alpha : \alpha \in \mathbb{R}, x + \alpha \mathbf{1}_e \in P_f\} \tag{15.24}$$

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

Matroids and Exchange

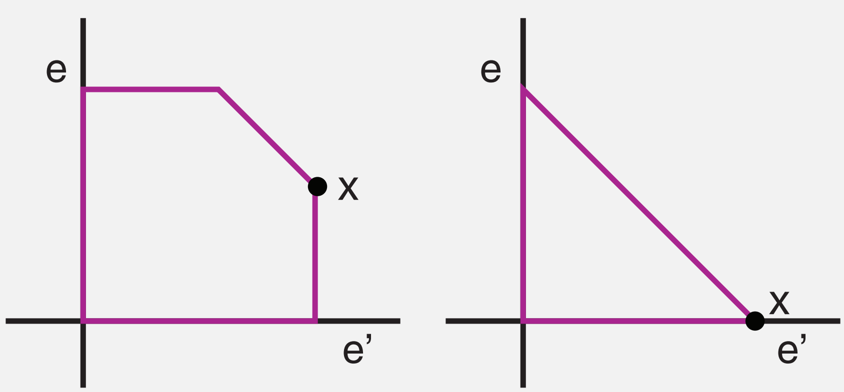
- Recall, matroids have a number of "exchange" properties.
- Also, recall that given a matroid $\mathcal{M} = (E, \mathcal{I})$, if $I \in \mathcal{I}$ is independent, and $e \in \text{span}(I)$, and $e' \in C(I, e)$ where $C(I, e)$ is the fundamental circuit created when adding e to I, then we have:

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

- 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 $e' \in \text{dep}(x, e) \setminus \{e\}$
- 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
$$\max\{\alpha : \alpha \in \mathbb{R}, x + \alpha(\mathbf{1}_e - \mathbf{1}_{e'}) \in P_f\} \tag{15.5}$$
- 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
$$\max\left\{\alpha : \alpha \in \mathbb{R}, (x + \alpha(\mathbf{1}_e - \mathbf{1}_{e'}))(A) \le f(A), \forall A \ni e, e' \notin A\right\} \tag{15.7}$$
- Which is identical to:
$$\max\left\{\alpha : \alpha \in \mathbb{R}, \alpha(\mathbf{1}_e - \mathbf{1}_{e'}))(A) \le f(A) - x(A), \forall A \supseteq \{e\}, e' \notin A\right\} \tag{15.8}$$

Exchange Capacity

- In such case, we get $\mathbf{1}_{e'}(A) = 0$, thus above identical to
$$\max\left\{\alpha : \alpha \in \mathbb{R}, \alpha \mathbf{1}_e(A) \le f(A) - x(A), \forall A \supseteq \{e\}, e' \notin A\right\} \tag{15.9}$$
- Restating, we've got
$$\max\left\{\alpha : \alpha \in \mathbb{R}, \alpha \le f(A) - x(A), \forall A \supseteq \{e\}, e' \notin A\right\} \tag{15.10}$$
- This max is achieved when
$$\alpha = \hat{c}(x; e, e') \stackrel{\text{def}}{=} \min\left\{f(A) - x(A), \forall A \supseteq \{e\}, e' \notin A\right\} \tag{15.11}$$
- $\hat{c}(x; e, e')$ 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 $x + \alpha(\mathbf{1}_e - \mathbf{1}_{e'}) \in P_f$.
- 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:

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

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:

$$\textit{rank}(0) = \max\left(y(E) : y \leq 0, y \in P_f\right) = \min\left(f(A) : A \subseteq E\right) \tag{15.7}$$

Min-Norm Point: Definition

- Restating what we saw before, we have:

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

- 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.
- x^* is called the minimum norm point of the base polytope.

Min-Norm Point: Examples