Submodular Functions, Optimization, and Applications to Machine Learning

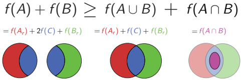
— Spring Quarter, Lecture 16 —

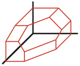
http://j.ee.washington.edu/~bilmes/classes/ee596b_spring_2014/

Prof. Jeff Bilmes

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

May 21st, 2014





Cumulative Outstanding Reading

- Good references for today: Schrijver-2003, Oxley-1992/2011, Welsh-1973, Goemans-2010, Cunningham-1984, Edmonds-1969, Choquet-1955, Grabisch/Marichal/Mesiar/Pap "Aggregation Functions", Lovász-1983, Bach-2011.
- 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 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,

$$\text{cl}(x) \stackrel{\text{def}}{=} \text{sat}(x) \triangleq \bigcup \{A : A \in \mathcal{D}(x)\} \tag{16.29}$$

$$= \bigcup \{A : A \subseteq E, x(A) = f(A)\} \tag{16.30}$$

$$= \{e : e \in E, \forall \alpha > 0, x + \alpha \mathbf{1}_e \notin P_f\} \tag{16.31}$$

- 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.,

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

$$= \{e' : \exists \alpha > 0, \text{ s.t. } x + \alpha(\mathbf{1}_e - \mathbf{1}_{e'}) \in P_f\} \tag{16.32}$$

A polymatroid function's polyhedron is a polymatroid.

Theorem 16.2.1

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

$$\textit{rank}(x) = \max\left(y(E) : y \leq x, y \in P_f\right) = \min\left(x(A) + f(E \setminus A) : A \subseteq E\right) \tag{16.5}$$

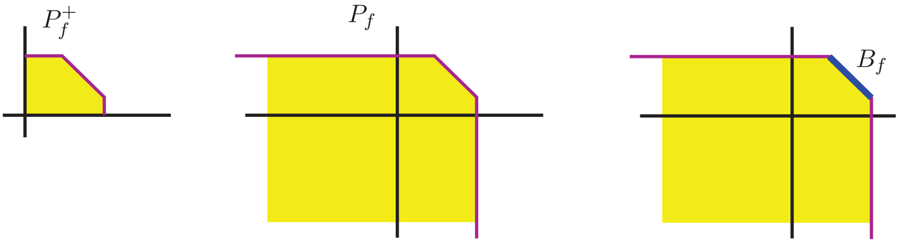
If we take x to be zero, we get:

Corollary 16.2.2

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

$$\textit{rank}(0) = \max\left(y(E) : y \leq 0, y \in P_f\right) = \min\left(f(A) : A \subseteq E\right) \tag{16.6}$$

Multiple Polytopes associated with f



$$P_f^+ = P_f \cap \left\{x \in \mathbb{R}^E : x \geq 0\right\} \tag{16.5}$$

$$P_f = \left\{x \in \mathbb{R}^E : x(S) \leq f(S), \forall S \subseteq E\right\} \tag{16.6}$$

$$B_f = P_f \cap \left\{x \in \mathbb{R}^E : x(E) = f(E)\right\} \tag{16.7}$$

Min-Norm Point: Definition

- Restating what we saw before, we have:

$$\max \{y(E) | y \in P_f, y \leq 0\} = \min \{f(X) | X \subseteq V\} \tag{16.12}$$

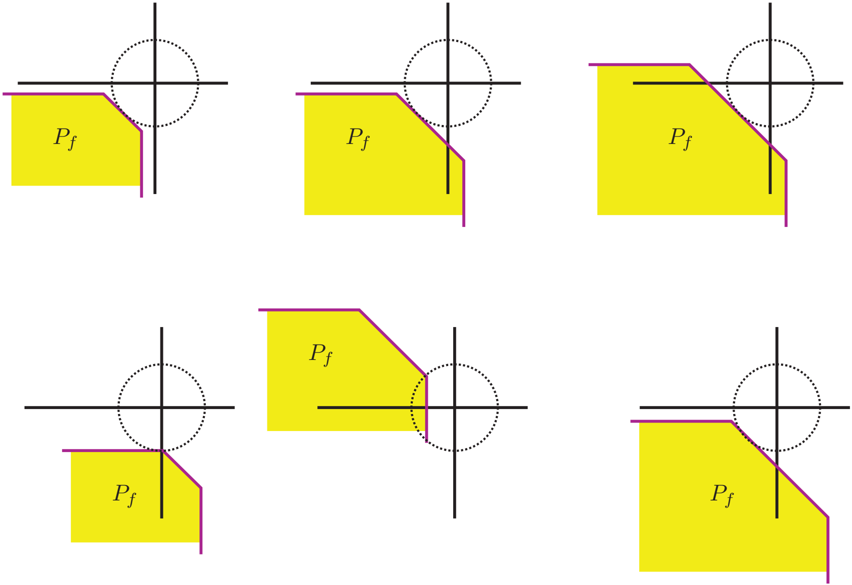
- Consider the optimization:

$$\begin{aligned} &\text{minimize} && \|x\|_2^2 && (16.13a) \\ &\text{subject to} && x \in B_f && (16.13b) \end{aligned}$$

  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



Min-Norm Point and Submodular Function Minimization

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

$$y^* = x^* \wedge 0 = (\min(x^*(e), 0) | e \in E) \tag{16.1}$$

$$A_- = \{e : x^*(e) < 0\} \tag{16.2}$$

$$A_0 = \{e : x^*(e) \leq 0\} \tag{16.3}$$

Min-Norm Point and Submodular Function Minimization

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

$$y^* = x^* \wedge 0 = (\min(x^*(e), 0) | e \in E) \tag{16.1}$$

$$A_- = \{e : x^*(e) < 0\} \tag{16.2}$$

$$A_0 = \{e : x^*(e) \leq 0\} \tag{16.3}$$

- Thus, we immediately have that:

$$A_- \subseteq A_0 \tag{16.4}$$

and that

$$x^*(A_-) = x^*(A_0) = y^*(A_-) = y^*(A_0) \tag{16.5}$$

Min-Norm Point and Submodular Function Minimization

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

$$y^* = x^* \wedge 0 = (\min(x^*(e), 0) | e \in E) \tag{16.1}$$

$$A_- = \{e : x^*(e) < 0\} \tag{16.2}$$

$$A_0 = \{e : x^*(e) \leq 0\} \tag{16.3}$$

- Thus, we immediately have that:

$$A_- \subseteq A_0 \tag{16.4}$$

  and that

$$x^*(A_-) = x^*(A_0) = y^*(A_-) = y^*(A_0) \tag{16.5}$$

- 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

$$y^* = x^* \wedge 0 = (\min(x^*(e), 0) | e \in E) \tag{16.1}$$
$$A_- = \{e : x^*(e) < 0\} \tag{16.2}$$
$$A_0 = \{e : x^*(e) \leq 0\} \tag{16.3}$$

- Thus, we immediately have that:

$$A_- \subseteq A_0 \tag{16.4}$$

and that

$$x^*(A_-) = x^*(A_0) = y^*(A_-) = y^*(A_0) \tag{16.5}$$

- 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 16.3.1

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 16.3.1

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 $\text{sat}(x^*) = E$. Thus, we can consider any $e \in E$ within $\text{dep}(x^*, e)$.
- Consider any pair (e, e') with $e' \in \text{dep}(x^*, e)$ 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 16.3.1

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 $\text{sat}(x^*) = E$. Thus, we can consider any $e \in E$ within $\text{dep}(x^*, e)$.
- Consider any pair (e, e') with $e' \in \text{dep}(x^*, e)$ 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

$$(x^* + \alpha \mathbf{1}_e - \alpha \mathbf{1}_{e'})(E) = x^*(E) + \alpha - \alpha = f(E) \tag{16.6}$$

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

...

Min-Norm Point and SFM

. . . proof of Thm. 16.3.1 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. 16.3.1 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 $\alpha > 0$,
$\Big(x^*(e)\Big)^2 + \Big(x^*(e')\Big)^2 < \Big(x^*_{\text{new}}(e)\Big)^2 + \Big(x^*_{\text{new}}(e')\Big)^2$

. . .

Min-Norm Point and SFM

. . . proof of Thm. 16.3.1 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. 16.3.1 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 $0 < \alpha < |x^*(e)|$ (Exercise:), again contradicting the optimality of x^*.

...

Min-Norm Point and SFM

...proof of Thm. 16.3.1 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 $0 < \alpha < |x^*(e)|$ (Exercise:), again contradicting the optimality of x^*.
- Thus, we must have $x^*(e') < 0$ (strict negativity).

...

Min-Norm Point and SFM

... proof of Thm. 16.3.1 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. 16.3.1 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. 16.3.1 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. 16.3.1 cont.

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

Min-Norm Point and SFM

. . . proof of Thm. 16.3.1 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. 16.3.1 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:

Min-Norm Point and SFM

... proof of Thm. 16.3.1 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:

$$x^*(A_-) = f(A_-) \tag{16.7}$$

Min-Norm Point and SFM

... proof of Thm. 16.3.1 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:

$$x^*(A_-) = f(A_-) \tag{16.7}$$
$$x^*(A_0) = f(A_0) \tag{16.8}$$

Min-Norm Point and SFM

. . . proof of Thm. 16.3.1 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:

$$x^*(A_-) = f(A_-) \tag{16.7}$$

$$x^*(A_0) = f(A_0) \tag{16.8}$$

$$x^*(A_-) = x^*(A_0) = y^*(E) = y^*(A_0) + \underbrace{y^*(E \setminus A_0)}_{=0} \tag{16.9}$$

Min-Norm Point and SFM

... proof of Thm. 16.3.1 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:

$$x^*(A_-) = f(A_-) \tag{16.7}$$

$$x^*(A_0) = f(A_0) \tag{16.8}$$

$$x^*(A_-) = x^*(A_0) = y^*(E) = y^*(A_0) + \underbrace{y^*(E \setminus A_0)}_{=0} \tag{16.9}$$

and therefore, all together we have

Min-Norm Point and SFM

. . . proof of Thm. 16.3.1 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:

$$x^*(A_-) = f(A_-) \tag{16.7}$$

$$x^*(A_0) = f(A_0) \tag{16.8}$$

$$x^*(A_-) = x^*(A_0) = y^*(E) = y^*(A_0) + \underbrace{y^*(E \setminus A_0)}_{=0} \tag{16.9}$$

and therefore, all together we have

$$f(A_-) = f(A_0) = x^*(A_-) = x^*(A_0) = y^*(E) \tag{16.10}$$

Min-Norm Point and SFM

... proof of Thm. 16.3.1 cont.

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

...

Min-Norm Point and SFM

. . . proof of Thm. 16.3.1 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. 16.3.1 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. 16.3.1 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. 16.3.1 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. 16.3.1 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. (16.7), 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. 16.3.1 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. (16.7), 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. 16.3.1 cont.

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

$$f(X) \geq x^*(X) > x^*(A_-) = f(A_-) \tag{16.11}$$

$\square$

Min-Norm Point and SFM

…proof of Thm. 16.3.1 cont.

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

$$f(X) \geq x^*(X) > x^*(A_-) = f(A_-) \tag{16.11}$$

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

$$f(X) \geq x^*(X) > x^*(A_0) = f(A_0) \tag{16.12}$$

□

Min-Norm Point and SFM

. . . proof of Thm. 16.3.1 cont.

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

$$f(X) \geq x^*(X) > x^*(A_-) = f(A_-) \tag{16.11}$$

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

$$f(X) \geq x^*(X) > x^*(A_0) = f(A_0) \tag{16.12}$$

- 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 16.3.2

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

$$A = A_- \cup \bigcup_{a \in A_m} \operatorname{dep}(x^*, a) \tag{16.13}$$

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

Min-norm point and other minimizers of f

proof of Thm. 16.3.2.

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

□

Min-norm point and other minimizers of f

proof of Thm. 16.3.2.

- 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. 16.3.2.

- 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^*.

□

Min-norm point and other minimizers of f

proof of Thm. 16.3.2.

- 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.

$\square$

Min-norm point and other minimizers of f

proof of Thm. 16.3.2.

- 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.
- Hence, for any $a \in A$, $\mathrm{dep}(x^*, a) \subseteq A$.

□

Min-norm point and other minimizers of f

proof of Thm. 16.3.2.

- 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.
- Hence, for any $a \in A$, $\mathrm{dep}(x^*, a) \subseteq A$.
- This means that $\bigcup_{a \in A} \mathrm{dep}(x^*, a) = A$.

$\square$

Min-norm point and other minimizers of f

proof of Thm. 16.3.2.

- 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

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

$\square$

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 $e \in A_0$.

Review

The next slide comes from lecture 12.

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 16.4.1

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

$$\begin{aligned} &\text{maximize} && w^\intercal x && (16.14a)\\ &\text{subject to} && x \in P_f && (16.14b) \end{aligned}$$

Optimization over P_f

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

$$\begin{aligned} &\text{maximize} && w^\intercal x && (16.14a)\\ &\text{subject to} && x \in P_f && (16.14b) \end{aligned}$$

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

$$\begin{aligned} &\text{maximize} && w^\intercal x && (16.14a) \\ &\text{subject to} && x \in P_f && (16.14b) \end{aligned}$$

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

Optimization over P_f

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

$$\begin{aligned} &\text{maximize} && w^\intercal x && (16.14a) \\ &\text{subject to} && x \in P_f && (16.14b) \end{aligned}$$

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

$$\begin{aligned} &\text{maximize} && w^\intercal x && (16.14a)\\ &\text{subject to} && x \in P_f && (16.14b)\end{aligned}$$

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

$$\begin{aligned} &\text{maximize} && w^\intercal x && (16.14a) \\ &\text{subject to} && x \in P_f && (16.14b) \end{aligned}$$

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

$$\begin{aligned} &\text{maximize} && w^\intercal x && (16.15a) \\ &\text{subject to} && x \in B_f && (16.15b) \end{aligned}$$

Optimization over P_f

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

$$\begin{aligned} &\text{maximize} && w^\intercal x && (16.14a)\\ &\text{subject to} && x \in P_f && (16.14b) \end{aligned}$$

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

$$\begin{aligned} &\text{maximize} && w^\intercal x && (16.15a)\\ &\text{subject to} && x \in B_f && (16.15b) \end{aligned}$$

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

A continuous extension of f

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

$$\begin{aligned} &\text{maximize} && w^\intercal x && (16.16a)\\ &\text{subject to} && x \in P_f && (16.16b) \end{aligned}$$

A continuous extension of f

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

$$\begin{aligned} \text{maximize} \quad & w^\intercal x && (16.16a) \\ \text{subject to} \quad & x \in P_f && (16.16b) \end{aligned}$$

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

$$\tilde{f}(w) = \max(wx : x \in P_f) \tag{16.17}$$

A continuous extension of f

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

$$\begin{aligned} \text{maximize} \quad & w^\intercal x && (16.16a) \\ \text{subject to} \quad & x \in P_f && (16.16b) \end{aligned}$$

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

$$\tilde{f}(w) = \max(wx : x \in P_f) \tag{16.17}$$

- 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 submodular 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

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

A continuous extension of submodular f

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

$$\tilde{f}(w) = \max(wx : x \in P_f) \tag{16.18}$$

A continuous extension of submodular f

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

$$\tilde{f}(w) = \max(wx : x \in P_f) \tag{16.18}$$

$$= \sum_{i=1}^{m} w(e_i) f(e_i | E_{i-1}) \tag{16.19}$$

A continuous extension of submodular f

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

$$\tilde{f}(w) = \max(wx : x \in P_f) \tag{16.18}$$

$$= \sum_{i=1}^{m} w(e_i) f(e_i | E_{i-1}) \tag{16.19}$$

$$= \sum_{i=1}^{m} w(e_i)(f(E_i) - f(E_{i-1})) \tag{16.20}$$

A continuous extension of submodular f

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

$$\tilde{f}(w) = \max(wx : x \in P_f) \tag{16.18}$$

$$= \sum_{i=1}^{m} w(e_i) f(e_i | E_{i-1}) \tag{16.19}$$

$$= \sum_{i=1}^{m} w(e_i)(f(E_i) - f(E_{i-1})) \tag{16.20}$$

$$= w(e_m) f(E_m) + \sum_{i=1}^{m-1} (w(e_i) - w(e_{i+1})) f(E_i) \tag{16.21}$$

A continuous extension of submodular 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

$$\tilde{f}(w) = \max(wx : x \in P_f) \tag{16.18}$$

$$= \sum_{i=1}^{m} w(e_i) f(e_i | E_{i-1}) \tag{16.19}$$

$$= \sum_{i=1}^{m} w(e_i)(f(E_i) - f(E_{i-1})) \tag{16.20}$$

$$= w(e_m) f(E_m) + \sum_{i=1}^{m-1} (w(e_i) - w(e_{i+1})) f(E_i) \tag{16.21}$$

- 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:

$$\tilde{f}(w) = \max(wx : x \in P_f) \tag{16.22}$$

A continuous extension of f

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

$$\tilde{f}(w) = \max(wx : x \in P_f) \tag{16.22}$$

- 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:

$$\tilde{f}(w) = \max(wx : x \in P_f) \tag{16.22}$$

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

$$\tilde{f}(w) = w(e_m)f(E_m) + \sum_{i=1}^{m-1} (w(e_i) - w(e_{i+1}))f(E_i) \tag{16.23}$$

A continuous extension of f

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

$$\tilde{f}(w) = \max(wx : x \in P_f) \tag{16.22}$$

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

$$\tilde{f}(w) = w(e_m)f(E_m) + \sum_{i=1}^{m-1} (w(e_i) - w(e_{i+1}))f(E_i) \tag{16.23}$$

$$= \sum_{i=1}^{m} \lambda_i f(E_i) \tag{16.24}$$

A continuous extension of f

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

$$\tilde{f}(w) = \max(wx : x \in P_f) \tag{16.22}$$

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

$$\tilde{f}(w) = w(e_m)f(E_m) + \sum_{i=1}^{m-1} (w(e_i) - w(e_{i+1}))f(E_i) \tag{16.23}$$

$$= \sum_{i=1}^{m} \lambda_i f(E_i) \tag{16.24}$$

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:

$$\tilde{f}(w) = \max(wx : x \in P_f) \tag{16.22}$$

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

$$\tilde{f}(w) = w(e_m)f(E_m) + \sum_{i=1}^{m-1}(w(e_i) - w(e_{i+1}))f(E_i) \tag{16.23}$$

$$= \sum_{i=1}^{m} \lambda_i f(E_i) \tag{16.24}$$

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

$$\begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} = \underbrace{(w_1 - w_2)}_{\lambda_1} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \underbrace{(w_2 - w_3)}_{\lambda_2} \begin{pmatrix} 1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \cdots + \underbrace{(w_{n-1} - w_n)}_{\lambda_{m-1}} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 0 \end{pmatrix} + \underbrace{(w_m)}_{\lambda_m} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{pmatrix} \tag{16.25}$$

An extension of f

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

$$\begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} = \underbrace{(w_1 - w_2)}_{\lambda_1} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \underbrace{(w_2 - w_3)}_{\lambda_2} \begin{pmatrix} 1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \cdots + \underbrace{(w_{n-1} - w_n)}_{\lambda_{m-1}} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 0 \end{pmatrix} + \underbrace{(w_m)}_{\lambda_m} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{pmatrix} \tag{16.25}$$

- 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 $i = 0, \ldots, n$

$$E_i \stackrel{\text{def}}{=} \{e_1, e_2, \ldots, e_i\} \tag{16.26}$$

An extension of f

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

$$E_i \stackrel{\text{def}}{=} \{e_1, e_2, \dots, e_i\} \tag{16.26}$$

- Note that

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

An extension of f

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

$$E_i \stackrel{\text{def}}{=} \{e_1, e_2, \dots, e_i\} \tag{16.26}$$

- Note that

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

- 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 $w(e_1) \geq w(e_2) \geq w(e_3) \geq \cdots \geq w(e_m)$.
- This means

$$w = (w(e_1), w(e_2), \ldots, w(e_m)) = (\underbrace{1, 1, 1, \ldots, 1}_{|A| \text{ times}}, \underbrace{0, 0, \ldots, 0}_{m-|A| \text{ times}}) \quad (16.27)$$

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

$$w = (w(e_1), w(e_2), \ldots, w(e_m)) = (\underbrace{1, 1, 1, \ldots, 1}_{|A| \text{ times}}, \underbrace{0, 0, \ldots, 0}_{m-|A| \text{ times}}) \quad (16.27)$$

  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

$$w = (w(e_1), w(e_2), \ldots, w(e_m)) = (\underbrace{1, 1, 1, \ldots, 1}_{|A| \text{ times}}, \underbrace{0, 0, \ldots, 0}_{m-|A| \text{ times}}) \quad (16.27)$$

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 $w(e_1) \geq w(e_2) \geq w(e_3) \geq \dots \geq w(e_m)$.
- This means

$$w = (w(e_1), w(e_2), \dots, w(e_m)) = (\underbrace{1, 1, 1, \dots, 1}_{|A| \text{ times}}, \underbrace{0, 0, \dots, 0}_{m-|A| \text{ times}}) \tag{16.27}$$

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

$$w = (w(e_1), w(e_2), \ldots, w(e_m)) = (\underbrace{1, 1, 1, \ldots, 1}_{|A| \text{ times}}, \underbrace{0, 0, \ldots, 0}_{m-|A| \text{ times}}) \tag{16.27}$$

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

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

$$= \mathbf{1}_A(m) f(E_m) + \sum_{i=1}^{m-1} (\mathbf{1}_A(i) - \mathbf{1}_A(i+1)) f(E_i) \tag{16.28}$$

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

$$w = (w(e_1), w(e_2), \ldots, w(e_m)) = (\underbrace{1, 1, 1, \ldots, 1}_{|A| \text{ times}}, \underbrace{0, 0, \ldots, 0}_{m-|A| \text{ times}}) \quad (16.27)$$

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

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

$$= \mathbf{1}_A(m) f(E_m) + \sum_{i=1}^{m-1} (\mathbf{1}_A(i) - \mathbf{1}_A(i+1)) f(E_i) \quad (16.28)$$

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

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

$$w = (w(e_1), w(e_2), \ldots, w(e_m)) = (\underbrace{1, 1, 1, \ldots, 1}_{|A| \text{ times}}, \underbrace{0, 0, \ldots, 0}_{m-|A| \text{ times}}) \quad (16.27)$$

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

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

$$= \mathbf{1}_A(m) f(E_m) + \sum_{i=1}^{m-1} (\mathbf{1}_A(i) - \mathbf{1}_A(i+1)) f(E_i) \quad (16.28)$$

$$= (\mathbf{1}_A(|A|) - \mathbf{1}_A(|A|+1)) f(E_{|A|}) = f(E_{|A|}) = f(A) \quad (16.29)$$

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

$$\tilde{f}(\mathbf{1}_A) = f(A), \tag{16.30}$$

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

$$\tilde{f}(\mathbf{1}_A) = f(A), \tag{16.30}$$

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

$$\begin{aligned} \tilde{f}(\mathbf{1}_A) &= \max\{\mathbf{1}_A x : x \in P_f\} && (16.31) \\ &= \max\{\mathbf{1}_A x : x(B) \le f(B), \forall B \subseteq E\} && (16.32) \\ & && (16.33) \end{aligned}$$

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

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

$$\tilde{f}(w) = \sum_{i=1}^{m} \lambda_i f(E_i) \tag{16.34}$$

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

$$\text{for } i \in \{1, \ldots, m\}, \quad \lambda_i = \begin{cases} w(e_i) - w(e_{i+1}) & \text{if } i < m \\ w(e_m) & \text{if } i = m \end{cases} \tag{16.35}$$

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

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

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

$$\tilde{f}(w) = \sum_{i=1}^{m} \lambda_i f(E_i) \tag{16.34}$$

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

$$\text{for } i \in \{1, \ldots, m\}, \quad \lambda_i = \begin{cases} w(e_i) - w(e_{i+1}) & \text{if } i < m \\ w(e_m) & \text{if } i = m \end{cases} \tag{16.35}$$

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

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

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

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

$$\tilde{f}(w) = \sum_{i=1}^{m} \lambda_i f(E_i) \tag{16.34}$$

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

$$\text{for } i \in \{1, \dots, m\}, \quad \lambda_i = \begin{cases} w(e_i) - w(e_{i+1}) & \text{if } i < m \\ w(e_m) & \text{if } i = m \end{cases} \tag{16.35}$$

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

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

Weighted gains vs. weighted functions

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

$$\tilde{f}(w) = \sum_{i=1}^{m} w(e_i) f(e_i | E_{i-1}) \tag{16.36}$$

$$= \sum_{i=1}^{m} w(e_i)(f(E_i) - f(E_{i-1})) \tag{16.37}$$

$$= w(e_m) f(E_m) + \sum_{i=1}^{m-1} (w(e_i) - w(e_{i+1})) f(E_i) \tag{16.38}$$

$$= \sum_{i=1}^{m} \lambda_i f(E_i) \tag{16.39}$$

Weighted gains vs. weighted functions

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

$$\tilde{f}(w) = \sum_{i=1}^{m} w(e_i) f(e_i | E_{i-1}) \tag{16.36}$$

$$= \sum_{i=1}^{m} w(e_i)(f(E_i) - f(E_{i-1})) \tag{16.37}$$

$$= w(e_m) f(E_m) + \sum_{i=1}^{m-1} (w(e_i) - w(e_{i+1})) f(E_i) \tag{16.38}$$

$$= \sum_{i=1}^{m} \lambda_i f(E_i) \tag{16.39}$$

- So $\tilde{f}(w)$ seen either as sum of weighted gain evaluatiosn (Eqn. (16.36), or as sum of weighted function evaluations (Eqn. (16.39)).

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

- Lovász showed that if a function $\tilde{f}(w)$ defined as in Eqn. (16.34) 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. (16.34) 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. (16.34) 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 16.4.1

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.(16.34) 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 16.4.1

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.(16.34) 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 16.4.1

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.(16.34) 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. 16.4.1 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. 16.4.1 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

$$\tilde{f}(\mathbf{1}_A + \mathbf{1}_B) = \tilde{f}(\mathbf{1}_{A\cup B} + \mathbf{1}_{A\cap B}) \tag{16.40}$$

$$= f(A \cup B) + f(A \cap B). \tag{16.41}$$

. . .

Lovász Extension, Submodularity and Convexity

. . . proof of Thm. 16.4.1 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

$$\tilde{f}(\mathbf{1}_A + \mathbf{1}_B) = \tilde{f}(\mathbf{1}_{A\cup B} + \mathbf{1}_{A\cap B}) \tag{16.40}$$

$$= f(A \cup B) + f(A \cap B). \tag{16.41}$$

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

$$w = (w(e_1), w(e_2), \dots, w(e_m)) \tag{16.42}$$

$$= (\underbrace{2, 2, \dots, 2}_{i \in C}, \underbrace{1, 1, \dots, 1}_{i \in A \triangle B}, \underbrace{0, 0, \dots, 0}_{i \in E \setminus (A \cup B)}) \tag{16.43}$$

. . .

Lovász Extension, Submodularity and Convexity

. . . proof of Thm. 16.4.1 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

$$\tilde{f}(\mathbf{1}_A + \mathbf{1}_B) = \tilde{f}(\mathbf{1}_{A\cup B} + \mathbf{1}_{A\cap B}) \tag{16.40}$$
$$= f(A \cup B) + f(A \cap B). \tag{16.41}$$

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

$$w = (w(e_1), w(e_2), \dots, w(e_m)) \tag{16.42}$$
$$= (\underbrace{2, 2, \dots, 2}_{i\in C}, \underbrace{1, 1, \dots, 1}_{i\in A\triangle B}, \underbrace{0, 0, \dots, 0}_{i\in E\setminus(A\cup B)}) \tag{16.43}$$

- 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. 16.4.1 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

$$\tilde{f}(\mathbf{1}_A + \mathbf{1}_B) = \tilde{f}(\mathbf{1}_{A\cup B} + \mathbf{1}_{A\cap B}) \tag{16.40}$$

$$= f(A \cup B) + f(A \cap B). \tag{16.41}$$

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

$$w = (w(e_1), w(e_2), \ldots, w(e_m)) \tag{16.42}$$

$$= (\underbrace{2, 2, \ldots, 2}_{i\in C}, \underbrace{1, 1, \ldots, 1}_{i\in A\triangle B}, \underbrace{0, 0, \ldots, 0}_{i\in E\setminus(A\cup B)}) \tag{16.43}$$

- 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, $\tilde{f}(w) = \tilde{f}(\mathbf{1}_A + \mathbf{1}_B) = f(A \cap B) + f(A \cup B)$.

. . .

Lovász Extension, Submodularity and Convexity

. . . proof of Thm. 16.4.1 cont.

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

$$0.5[f(A \cap B) + f(A \cup B)] \tag{16.47}$$

□

Lovász Extension, Submodularity and Convexity

. . . proof of Thm. 16.4.1 cont.

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

$$0.5[f(A \cap B) + f(A \cup B)] = 0.5[\tilde{f}(\mathbf{1}_A + \mathbf{1}_B)] \tag{16.44}$$

(16.47)

Lovász Extension, Submodularity and Convexity

. . . proof of Thm. 16.4.1 cont.

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

$$0.5[f(A \cap B) + f(A \cup B)] = 0.5[\tilde{f}(\mathbf{1}_A + \mathbf{1}_B)] \tag{16.44}$$

$$= \tilde{f}(0.5\mathbf{1}_A + 0.5\mathbf{1}_B) \tag{16.45}$$

(16.47)

□

Lovász Extension, Submodularity and Convexity

. . . proof of Thm. 16.4.1 cont.

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

$$0.5[f(A \cap B) + f(A \cup B)] = 0.5[\tilde{f}(\mathbf{1}_A + \mathbf{1}_B)] \quad (16.44)$$

$$= \tilde{f}(0.5\mathbf{1}_A + 0.5\mathbf{1}_B) \quad (16.45)$$

$$\leq 0.5\tilde{f}(\mathbf{1}_A) + 0.5\tilde{f}(\mathbf{1}_B) \quad (16.46)$$

$$(16.47)$$

□

Lovász Extension, Submodularity and Convexity

. . . proof of Thm. 16.4.1 cont.

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

$$0.5[f(A \cap B) + f(A \cup B)] = 0.5[\tilde{f}(\mathbf{1}_A + \mathbf{1}_B)] \tag{16.44}$$

$$= \tilde{f}(0.5\mathbf{1}_A + 0.5\mathbf{1}_B) \tag{16.45}$$

$$\leq 0.5\tilde{f}(\mathbf{1}_A) + 0.5\tilde{f}(\mathbf{1}_B) \tag{16.46}$$

$$= 0.5(f(A) + f(B)) \tag{16.47}$$

□

Lovász Extension, Submodularity and Convexity

. . . proof of Thm. 16.4.1 cont.

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

$$0.5[f(A \cap B) + f(A \cup B)] = 0.5[\tilde{f}(\mathbf{1}_A + \mathbf{1}_B)] \tag{16.44}$$

$$= \tilde{f}(0.5\mathbf{1}_A + 0.5\mathbf{1}_B) \tag{16.45}$$

$$\leq 0.5\tilde{f}(\mathbf{1}_A) + 0.5\tilde{f}(\mathbf{1}_B) \tag{16.46}$$

$$= 0.5(f(A) + f(B)) \tag{16.47}$$

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

$$f(A \cup B) + f(A \cap B) \leq f(A) + f(B) \tag{16.48}$$

so f must be submodular.

$\square$

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

Lovász - Submodularity - 1983

Submodular functions and convexity

L. Lovász
Eötvös Loránd University, Department of Analysis I, Múzeum krt. 6–8, H-1088 Budapest, Hungary

0. Introduction

In "continuous" optimization convex functions play a central role. Besides elementary tools like differentiation, various methods for finding the minimum of a convex function constitute the main body of nonlinear optimization. But even linear programming may be viewed as the optimization of very special (linear) objective functions over very special convex domains (polyhedra). There are several reasons for this popularity of convex functions:

– Convex functions occur in many mathematical models in economy, engineering, and other sciencies. Convexity is a very natural property of various functions and domains occuring in such models; quite often the only non-trivial property which can be stated in general.