Submodular Functions, Optimization, and Applications to Machine Learning

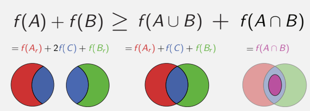
— Spring Quarter, Lecture 16 —

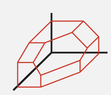
http://j.ee.washington.edu/~bilmes/classes/ee596b_spring_2014/

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 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,

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

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

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

- 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.,

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

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

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:

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

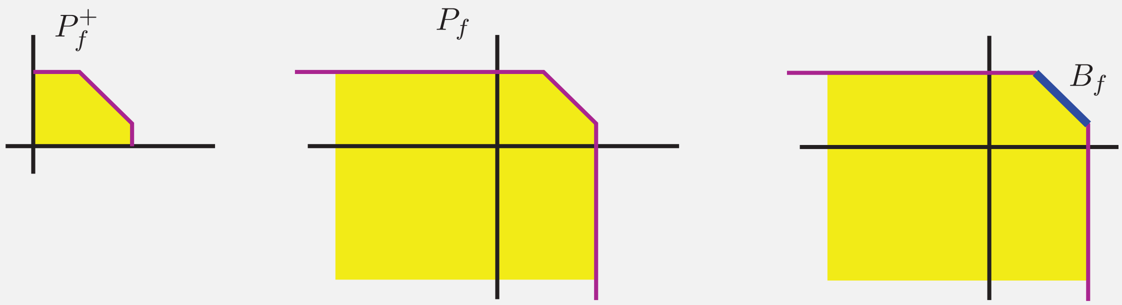
If we take x to be zero, we get:

Corollary 16.2.2

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

$$\textit{rank}(0) = \max\left(y(E) : y \leq 0, y \in P_f\right) = \min\left(f(A) : A \subseteq E\right) \tag{16.7}$$

Multiple Polytopes associated with f



$$P_f^+ = P_f \cap \left\{x \in \mathbb{R}^E : x \geq 0\right\} \tag{16.6}$$

$$P_f = \left\{x \in \mathbb{R}^E : x(S) \leq f(S), \forall S \subseteq E\right\} \tag{16.7}$$

$$B_f = P_f \cap \left\{x \in \mathbb{R}^E : x(E) = f(E)\right\} \tag{16.8}$$

Min-Norm Point: Definition

- Restating what we saw before, we have:

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

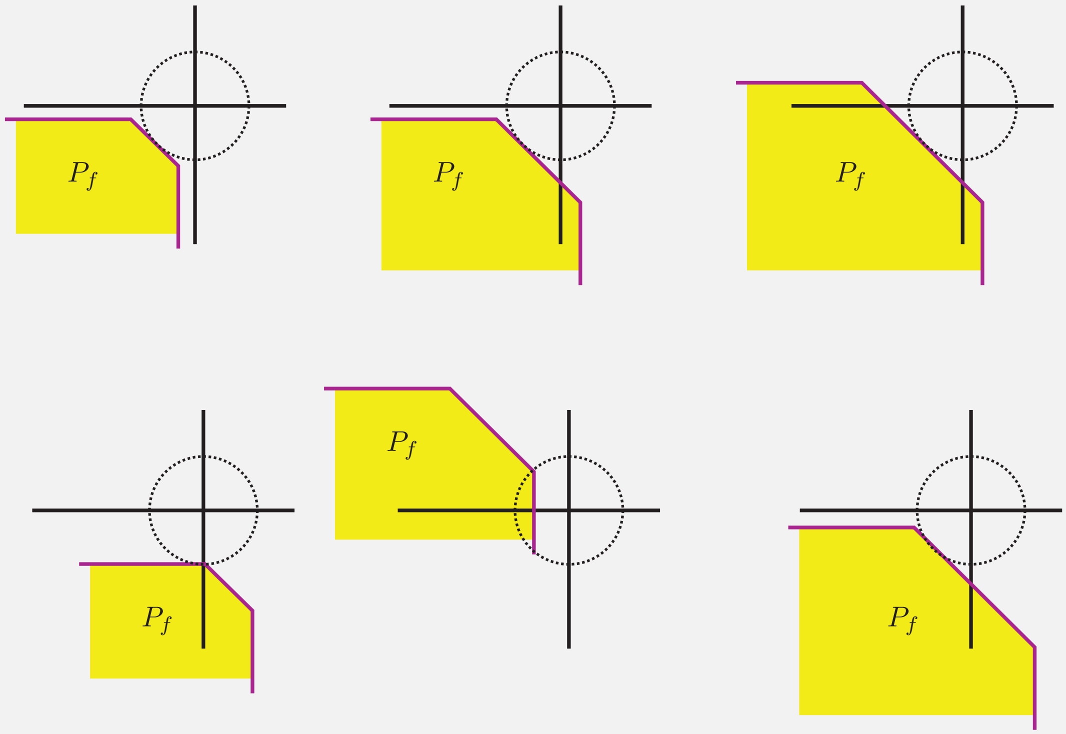
- Consider the optimization:

$$\text{minimize} \quad \|x\|_2^2 \tag{16.13a}$$

$$\text{subject to} \quad x \in B_f \tag{16.13b}$$

  where B_f is the base polytope of submodular f, and $\|x\|_2^2 = \sum_{e \in E} x(e)^2$ is the squared 2-norm. Let x^* be the optimal solution.
- Note, x^* is the unique optimal solution since we have a strictly convex objective over a set of convex constraints.
- x^* is called the minimum norm point of the base polytope.

Min-Norm Point: Examples



Min-Norm Point and Submodular Function Minimization

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

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

- Thus, we immediately have that:

$$x^*(A_-) = x^*(A_0) = y^*(A_-) = y^*(A_0) \tag{16.4}$$

- 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 $\text{sat}(x^*) = E$. 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

$$(x^* + \alpha \mathbf{1}_e - \alpha \mathbf{1}_{e'})(E) = x^*(E) + \alpha - \alpha = f(E) \tag{16.5}$$

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

$$x^*(A_-) = f(A_-) \tag{16.6}$$
$$x^*(A_0) = f(A_0) \tag{16.7}$$
$$x^*(A_-) = x^*(A_0) = y^*(E) \tag{16.8}$$

and therefore, all together we have

$$f(A_-) = f(A_0) = x^*(A_-) = x^*(A_0) = y^*(E) \tag{16.9}$$

. . .

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. (16.10), 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

$$f(X) \geq x^*(X) > x^*(A_-) = f(A_-) \tag{16.10}$$

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

$$f(X) \geq x^*(X) > x^*(A_0) = f(A_0) \tag{16.11}$$

- 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.
- 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.
- 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:

$$A = A_- \cup \bigcup_{a \in A_m} \mathrm{dep}(x^*, a) \tag{16.12}$$

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.
- 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

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

□

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

Optimization over P_f

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

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

- 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 $w \in \mathbb{R}^E_+$,

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

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

A continuous extension of f

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

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

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

$$\tilde{f}(w) = \max(wx : x \in P_f) \qquad (16.16)$$

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

A continuous extension of f

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

$$\tilde{f}(w) = \max(wx : x \in P_f) \tag{16.17}$$
$$= \sum_{i=1}^{m} w(e_i) f(e_i | E_{i-1}) \tag{16.18}$$
$$= \sum_{i=1}^{m} w(e_i)(f(E_i) - f(E_{i-1})) \tag{16.19}$$
$$= w(e_m) f(E_m) + \sum_{i=1}^{m-1} (w(e_i) - w(e_{i+1})) f(E_i) \tag{16.20}$$

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

A continuous extension of f

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

$$\tilde{f}(w) = \max(wx : x \in P_f) \tag{16.21}$$

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

$$\tilde{f}(w) = w(e_m) f(E_m) + \sum_{i=1}^{m-1} (w(e_i) - w(e_{i+1})) f(E_i) \tag{16.22}$$
$$= \sum_{i=1}^{m} \lambda_i f(E_i) \tag{16.23}$$

  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

$$\begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} = \underbrace{(w_1 - w_2)}_{\lambda_1} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \underbrace{(w_2 - w_3)}_{\lambda_2} \begin{pmatrix} 1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \cdots + \underbrace{(w_{n-1} - w_n)}_{\lambda_{m-1}} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 0 \end{pmatrix} + \underbrace{(w_m)}_{\lambda_m} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{pmatrix} \tag{16.24}$$

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

An extension of f

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

$$E_i \stackrel{\text{def}}{=} \{e_1, e_2, \dots, e_i\} \tag{16.25}$$

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

$$w = (w(e_1), w(e_2), \dots, w(e_m)) = (\underbrace{1, 1, 1, \dots, 1}_{|A| \text{ times}}, \underbrace{0, 0, \dots, 0}_{m-|A| \text{ times}}) \tag{16.26}$$

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

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

$$= \mathbf{1}_A(m) f(E_m) + \sum_{i=1}^{m-1} (\mathbf{1}_A(i) - \mathbf{1}_A(i+1)) f(E_i) \tag{16.27}$$

$$= (\mathbf{1}_A(|A|) - \mathbf{1}_A(|A|+1)) f(E_{|A|}) = f(E_{|A|}) = f(A) \tag{16.28}$$

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

$$\tilde{f}(\mathbf{1}_A) = f(A), \tag{16.29}$$

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

$$\tilde{f}(\mathbf{1}_A) = \max \{\mathbf{1}_A x : x \in P_f\} \tag{16.30}$$

$$= \max \{\mathbf{1}_A x : x(B) \leq f(B), \forall B \subseteq E\} \tag{16.31}$$

(16.32)

An extension of f

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

$$\tilde{f}(w) = \sum_{i=1}^{m} \lambda_i f(E_i) \tag{16.33}$$

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

$$\text{for } i \in \{1, \dots, m\}, \quad \lambda_i = \begin{cases} w(e_i) - w(e_{i+1}) & \text{if } i < m \\ w(e_m) & \text{if } i = m \end{cases} \tag{16.34}$$

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

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

Weighted gains vs. weighted functions

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

$$\tilde{f}(w) = \sum_{i=1}^{m} w(e_i) f(e_i | E_{i-1}) \tag{16.35}$$

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

$$= w(e_m) f(E_m) + \sum_{i=1}^{m-1} (w(e_i) - w(e_{i+1})) f(E_i) \tag{16.37}$$

$$= \sum_{i=1}^{m} \lambda_i f(E_i) \tag{16.38}$$

- So $\tilde{f}(w)$ seen either as sum of weighted gain evaluatiosn (Eqn. (16.35), or as sum of weighted function evaluations (Eqn. (16.38)).

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

- Lovász showed that if a function $\tilde{f}(w)$ defined as in Eqn. (16.33) 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.(16.33) 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 $A \subseteq E$.
- Now, given $A, B \subseteq E$, we will show that

$$\tilde{f}(\mathbf{1}_A + \mathbf{1}_B) = \tilde{f}(\mathbf{1}_{A\cup B} + \mathbf{1}_{A\cap B}) \tag{16.39}$$

$$= f(A \cup B) + f(A \cap B). \tag{16.40}$$

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

$$w = (w(e_1), w(e_2), \dots, w(e_m)) \tag{16.41}$$

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

- 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 $A, B \subseteq E$,

$$0.5[f(A \cap B) + f(A \cup B)] = 0.5[\tilde{f}(\mathbf{1}_A + \mathbf{1}_B)] \tag{16.43}$$

$$= \tilde{f}(0.5\mathbf{1}_A + 0.5\mathbf{1}_B) \tag{16.44}$$

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

$$= 0.5(f(A) + f(B)) \tag{16.46}$$

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

$$f(A \cup B) + f(A \cap B) \leq f(A) + f(B) \tag{16.47}$$

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 sciences. 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.

Integration and Aggregation

- Integration is just summation (e.g., the $\int$ symbol has as its origins a sum).
- Lebesgue integration allows integration w.r.t. an underlying measure μ of sets. E.g., given measurable function f, we can define

$$\int_X f du = \sup I_X(s) \tag{16.48}$$

where $I_X(s) = \sum_{i=1}^n c_i \mu(X \cap X_i)$, and where we take the sup over all measurable functions s such that $0 \leq s \leq f$ and $s(x) = \sum_{i=1}^n c_i I_{X_i}(x)$ and where $I_{X_i}(x)$ is indicator of membership of set X_i, with $c_i > 0$.

Integration, Aggregation, and Weighted Averages

- In finite discrete spaces, Lebesgue integration is just a weighted average, and can be seen as an aggregation function.
- I.e., given a weight vector $w \in [0,1]^E$ for some finite ground set E, then for any $x \in \mathbb{R}^E$ we have

$$\text{WAVG}(x) = \sum_{e \in E} x(e)w(e) \tag{16.49}$$

- Consider $\mathbf{1}_e$ for $e \in E$, we have

$$\text{WAVG}(\mathbf{1}_e) = w(e) \tag{16.50}$$

so seen as a function on the hypercube vertices, the entire WAVG function is given based on values on a subset of the vertices of this hypercube, i.e., $\{\mathbf{1}_e : e \in E\}$. Moreover, we are interpolating as in

$$\text{WAVG}(x) = \sum_{e \in E} x(e)w(e) = \sum_{e \in E} x(e)\text{WAVG}(\mathbf{1}_e) \tag{16.51}$$

- Note, WAVG function is linear in the weights w.

Integration, Aggregation, and Weighted Averages

- More complex "nonlinear" aggregation functions can be constructed by defining the aggregation function on all vertices of the hypercube. I.e., for each $\mathbf{1}_A : A \subseteq E$ we might have (for all $A \subseteq E$):

$$\text{AG}(\mathbf{1}_A) = w_A \tag{16.52}$$

- What then might AG(x) be for some $x \in \mathbb{R}^E$? Our weighted average functions might look something more like the r.h.s. in:

$$\text{AG}(x) = \sum_{A \subseteq E} x(A) w_A = \sum_{A \subseteq E} x(A)\text{AG}(\mathbf{1}_A) \tag{16.53}$$

- Note, we can define $w(e) = w'(e)$ and $w(A) = 0, \forall A : |A| > 1$ and get back previous (normal) weighted average, in that

$$\text{WAVG}_{w'}(x) = \text{AG}_w(x) \tag{16.54}$$

- Set function $f : 2^E \to \mathbb{R}$ is a game if f is normalized $f(\emptyset) = 0$.

Integration, Aggregation, and Weighted Averages

- Set function $f : 2^E \to \mathbb{R}$ is called a capacity if it is monotone non-decreasing, i.e., $f(A) \leq f(B)$ whenever $A \subseteq B$.
- A Boolean function f is any function $f : \{0,1\}^m \to \{0,1\}$ and is a pseudo-Boolean function if $f : \{0,1\}^m \to \mathbb{R}$.
- Any set function corresponds to a pseudo-Boolean function. I.e., given $f : 2^E \to \mathbb{R}$, form $f_b : \{0,1\}^m \to \mathbb{R}$ as $f_b(x) = f(A_x)$ where the A, x bijection is $A = \{e \in E : x_e = 1\}$ and $x = \mathbf{1}_A$.
- Also, If we have an expression for f_b we can construct a set function f as $f(A) = f_b(\mathbf{1}_A)$. We can also often relax f_b to any $x \in [0,1]^m$.
- We saw this for Lovász extension.
- It turns out that a concept essentially identical to the Lovász extension was derived much earlier, in 1954, and using this derivation (via integration) leads to deeper intuition.

Choquet integral

Definition 16.5.1

Let f be any capacity on E and $w \in \mathbb{R}_+^E$. The Choquet integral (1954) of w w.r.t. f is defined by

$$C_f(w) = \sum_{i=1}^{m} (w_{e_i} - w_{e_{i+1}}) f(E_i) \tag{16.55}$$

where in the sum, we have sorted and renamed the elements of E so that $w_{e_1} \geq w_{e_2} \geq \cdots \geq w_{e_m} \geq w_{e_{m+1}} = 0$, and where $E_i = \{e_1, e_2, \ldots, e_i\}$.

- We immediately see that an equivalent formula is as follows:

$$C_f(w) = \sum_{i=1}^{m} w(e_i)(f(E_i) - f(E_{i-1})) \tag{16.56}$$

where $E_0 \stackrel{\text{def}}{=} \emptyset$.

Min-Norm Point and SFM Lovász extension Choquet Integration Lovász extn., defs/props

Choquet integral

Definition 16.5.1

Let f be any capacity on E and $w \in \mathbb{R}_+^E$. The Choquet integral (1954) of w w.r.t. f is defined by

$$C_f(w) = \sum_{i=1}^{m} (w_{e_i} - w_{e_{i+1}}) f(E_i) \tag{16.55}$$

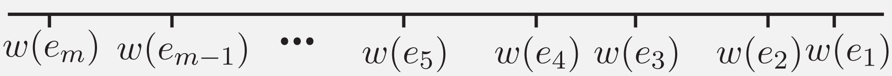
where in the sum, we have sorted and renamed the elements of E so that $w_{e_1} \geq w_{e_2} \geq \cdots \geq w_{e_m} \geq w_{e_{m+1}} = 0$, and where $E_i = \{e_1, e_2, \ldots, e_i\}$.

- BTW: this again essentially Abel's partial summation formula: Given two arbitrary sequences $\{a_n\}$ and $\{b_n\}$ with $A_n = \sum_{k=1}^{n} a_k$, we have

$$\sum_{k=m}^{n} a_k b_k = \sum_{k=m}^{n} A_k(b_k - b_{k+1}) + A_n b_{n+1} - A_{m-1} b_m \tag{16.57}$$

The "integral" in the Choquet integral

- Thought of as an integral over $\mathbb{R}$ of a piece-wise constant function.
- First note, assuming E is ordered according to descending w, so that $w(e_1) \geq w(e_2) \geq \dots \geq w(e_{m-1}) \geq w(e_m)$, then $E_i = \{e_1, e_2, \dots, e_i\} = \{e \in E : w_e \geq w_{e_i}\}$.
- For any $w_{e_i} > \alpha \geq w_{e_{i+1}}$ we also have $E_i = \{e_1, e_2, \dots, e_i\} = \{e \in E : w_e > \alpha\}$.
- Consider segmenting the real-axis at boundary points w_{e_i}, right most is w_{e_1}.



- A function can be defined on a segment of $\mathbb{R}$, namely $w_{e_i} > \alpha \geq w_{e_{i+1}}$. This function $F_i : [w_{e_{i+1}}, w_{e_i}) \to \mathbb{R}$ is defined as

$$F_i(\alpha) = f(\{e \in E : w_e > \alpha\}) = f(E_i) \tag{16.58}$$

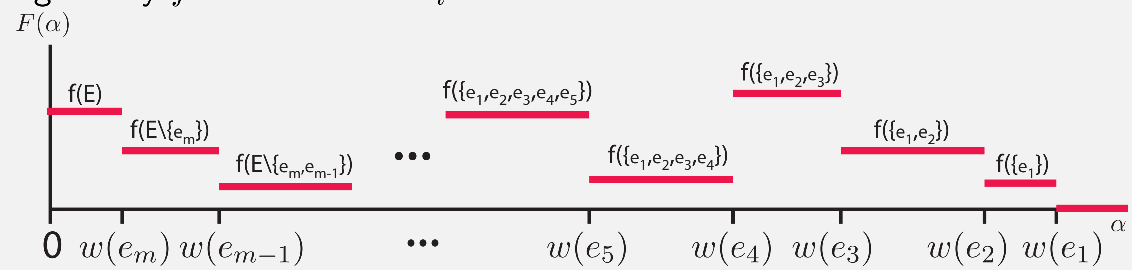
Min-Norm Point and SFM | Lovász extension | Choquet Integration | Lovász extn., defs/props

The "integral" in the Choquet integral

- We can generalize this to multiple segments of $\mathbb{R}$ (for now, take $w \in \mathbb{R}_+^E$). The piecewise-constant function is defined as:

$$F(\alpha) = \begin{cases} f(E) & \text{if } 0 \leq \alpha < w_m \\ f(\{e \in E : w_e > \alpha\}) & \text{if } w_{e_{i+1}} \leq \alpha < w_{e_i},\ i \in \{1, \dots, m-1\} \\ 0 & \text{if } w_1 < \alpha \end{cases}$$

- Visualizing a piecewise constant function, where the constant values are given by f evaluated on E_i for each i



Note, what is depicted may be a game but not a capacity.

The "integral" in the Choquet integral

- Now consider the integral, with $w \in \mathbb{R}_+^E$, and normalized f so that $f(\emptyset) = 0$. Recall $w_{m+1} \stackrel{\text{def}}{=} 0$.

$$\tilde{f}(w) \stackrel{\text{def}}{=} \int_0^\infty F(\alpha) d\alpha \tag{16.59}$$

$$= \int_0^\infty f(\{e \in E : w_e > \alpha\}) d\alpha \tag{16.60}$$

$$= \int_{w_{m+1}}^\infty f(\{e \in E : w_e > \alpha\}) d\alpha \tag{16.61}$$

$$= \sum_{i=1}^m \int_{w_{i+1}}^{w_i} f(\{e \in E : w_e > \alpha\}) d\alpha \tag{16.62}$$

$$= \sum_{i=1}^m \int_{w_{i+1}}^{w_i} f(E_i) d\alpha = \sum_{i=1}^m f(E_i)(w_i - w_{i+1}) \tag{16.63}$$

The "integral" in the Choquet integral

- But we saw before that $\sum_{i=1}^m f(E_i)(w_i - w_{i+1})$ is just the Lovász extension of a function f.
- Thus, we have the following definition:

Definition 16.5.2

Given $w \in \mathbb{R}_+^E$, the Lovász extension (equivalently Choquet integral) may be defined as follows:

$$\tilde{f}(w) \stackrel{\text{def}}{=} \int_0^\infty F(\alpha) d\alpha \tag{16.64}$$

where the function F is defined as before.

- Note that it is not necessary in general to require $w \in \mathbb{R}_+^E$ (i.e., we can take $w \in \mathbb{R}^E$) nor that f be non-negative, but it is a bit more involved. Above is the simple case.

Choquet integral and aggregation

- Recall, we want to produce some notion of generalized aggregation function having the flavor of:

$$\mathsf{AG}(x) = \sum_{A \subseteq E} x(A) w_A = \sum_{A \subseteq E} x(A)\mathsf{AG}(\mathbf{1}_A) \tag{16.65}$$

how does this correspond to Lovász extension?
- Let us partition the hypercube $[0,1]^m$ into q polytopes, each defined by a set of vertices $\mathcal{V}_1, \mathcal{V}_2, \dots, \mathcal{V}_q$.
- E.g., for each i, $\mathcal{V}_i = \{\mathbf{1}_{A_1}, \mathbf{1}_{A_2}, \dots, \mathbf{1}_{A_k}\}$ (k vertices) and the convex hull of V_i defines the i^{th} polytope.
- This forms a "triangulation" of the hypercube.
- For any $x \in [0,1]^m$ there is a (not necessarily unique) $\mathcal{V}(x) = \mathcal{V}_j$ for some j such that $x \in \text{conv}(\mathcal{V}(x))$.

Choquet integral and aggregation

- For $x \in [0,1]^m$, let us define the (unique) coefficients $\alpha_0^x(A)$ and $\alpha_i^x(A)$ so that x can be represented as a weighted combination of vertices of $\mathcal{V}(x)$. Note that many of these coefficient are often zero.
- From this, we can define an aggregation function of the form

$$\mathsf{AG}(x) \stackrel{\text{def}}{=} \sum_{A:\mathbf{1}_A \in \mathcal{V}(x)} \Big(\alpha_0^x(A) + \sum_{i=1}^{m} \alpha_i^x(A) x_i\Big)\mathsf{AG}(\mathbf{1}_A) \tag{16.66}$$

Choquet integral and aggregation

- We can define a canonical triangulation of the hypercube in terms of permutations of the coordinates. I.e., given some permutation σ, define

$$\operatorname{conv}(\mathcal{V}_\sigma) = \left\{x \in [0,1]^n | x_{\sigma(1)} \geq x_{\sigma(2)} \geq \dots \geq x_{\sigma(m)}\right\} \tag{16.67}$$

Then these $m!$ blocks of the partition are called the canonical partitions of the hypercube.

- With this, we can define $\{\mathcal{V}_i\}_i$ as the vertices of $\operatorname{conv}(\mathcal{V}_\sigma)$ for each permutation σ.
- In this case, we have:

Proposition 16.5.3

The above linear interpolation in Eqn. (16.66) *using the canonical partition yields the Lovász extension.*

- So the Lovász extension can be seen as a generalized aggregation function.

Lovász extension, defined in multiple ways

- As shorthand notation, lets use $\{w \geq \alpha\} \equiv \{e \in E : w(e) \geq \alpha\}$, called the weak α-sup-level set of w. A similar definition holds for $\{w > \alpha\}$ (called the strong α-sup-level set of w).
- Given any $w \in \mathbb{R}^E$, sort E as $w(e_1) \geq w(e_2) \geq \dots \geq w(e_m)$. Also, w.l.o.g., number elements of w so that $w_1 \geq w_2 \geq \dots \geq w_m$.
- We have already seen how we can define the Lovász extension for any (not necessarily submodular) function f in the following equivalent ways:

$$\tilde{f}(w) = \sum_{i=1}^{m} w(e_i) f(e_i|E_{i-1}) \tag{16.68}$$

$$= \sum_{i=1}^{m-1} f(E_i)(w(e_i) - w(e_{i+1})) + f(E)w(e_m)a \tag{16.69}$$

$$= \sum_{i=1}^{m-1} \lambda_i f(E_i) \tag{16.70}$$

Lovász extension, as integral

- Additional ways we can define the Lovász extension for any (not necessarily submodular) but normalized function f include:

$$\tilde{f}(w) = \sum_{i=1}^{m} w(e_i) f(e_i | E_{i-1}) \tag{16.71}$$

$$= \sum_{i=1}^{m-1} f(E_i)(w(e_i) - w(e_{i+1})) + f(E) w(e_m) \tag{16.72}$$

$$= \int_{\min\{w_1,\dots,w_m\}}^{+\infty} f(\{w \geq \alpha\}) d\alpha + f(E) \min\{w_1, \dots, w_m\} \tag{16.73}$$

$$= \int_{0}^{+\infty} f(\{w \geq \alpha\}) d\alpha + \int_{-\infty}^{0} [f(\{w \geq \alpha\}) - f(E)] d\alpha \tag{16.74}$$

general Lovász extension, as simple integral

- In fact, we have that, given function f, and any $w \in \mathbb{R}^E$:

$$\tilde{f}(w) = \int_{-\infty}^{+\infty} \hat{f}(\alpha) d\alpha \tag{16.75}$$

where

$$\hat{f}(\alpha) = \begin{cases} f(\{w \geq \alpha\}) & \text{if } \alpha >= 0 \\ f(\{w \geq \alpha\}) - f(E) & \text{if } \alpha < 0 \end{cases} \tag{16.76}$$

- So we can write it as a simple integral over the right function.
- These make it easier to see certain properties of the Lovász extension. But first, we show the above.

Lovász extension, as integral

- To show Eqn. (16.73), first note that the r.h.s. terms are the same since $w(e_m) = \min\{w_1, \dots, w_m\}$.
- Then, consider that, as a function of α, we have

$$f(\{w \geq \alpha\}) = \begin{cases} 0 & \text{if } \alpha > w(e_1) \\ f(E_k) & \text{if } \alpha \in (w(e_{k+1}), w(e_k)), k \in \{1, \dots, m-1\} \\ f(E) & \text{if } \alpha < w(e_m) \end{cases} \tag{16.77}$$

  we use open intervals since sets of zero measure don't change integration.
- Inside the integral, then, this recovers Eqn. (16.72).

Lovász extension, as integral

- To show Eqn. (16.74), start w. Eqn. (16.73), note $w_m = \min\{w_1, \dots, w_m\}$, take any $\beta \leq \min\{0, w_1, \dots, w_m\}$, and form:

$$\begin{aligned}
\tilde{f}(w) &= \int_{w_m}^{+\infty} f(\{w \geq \alpha\}) d\alpha + f(E) \min\{w_1, \dots, w_m\} \\
&= \int_{\beta}^{+\infty} f(\{w \geq \alpha\}) d\alpha - \int_{\beta}^{w_m} f(\{w \geq \alpha\}) d\alpha + f(E) \int_0^{w_m} d\alpha \\
&= \int_{\beta}^{+\infty} f(\{w \geq \alpha\}) d\alpha - \int_{\beta}^{w_m} f(E) d\alpha + \int_0^{w_m} f(E) d\alpha \\
&= \int_{0}^{+\infty} f(\{w \geq \alpha\}) d\alpha + \int_{\beta}^{0} f(\{w \geq \alpha\}) d\alpha - \int_{\beta}^{0} f(E) d\alpha \\
&= \int_{0}^{+\infty} f(\{w \geq \alpha\}) d\alpha + \int_{\beta}^{0} [f(\{w \geq \alpha\}) - f(E)] d\alpha
\end{aligned}$$

  and then let $\beta \to \infty$ and we get.

$$= \int_{0}^{+\infty} f(\{w \geq \alpha\}) d\alpha + \int_{-\infty}^{0} [f(\{w \geq \alpha\}) - f(E)] d\alpha$$

Lovász extension properties

- Using the above, have the following (some of which we've seen):

Theorem 16.6.1

Let $f, g : 2^E \to \mathbb{R}$ be normalized ($f(\emptyset) = g(\emptyset) = 0$). Then

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

Lovász extension properties: ex. property 3

- Consider property property 3, for example, which says that $\tilde{f}(w + \alpha \mathbf{1}_E) = \tilde{f}(w) + \alpha f(E)$.
- This means that, say when $m = 2$, that as we move along the line $w_1 = w_2$, the Lovász extension scales linearly.
- And if $f(E) = 0$, then the Lovász extension is constant along the direction $\mathbf{1}_E$.

Lovász extension properties

- Given Eqns. (16.71) through (16.74), most of the above properties are relatively easy to derive.
- For example, if f is symmetric, and since $f(E) = f(\emptyset) = 0$, we have

$$\tilde{f}(-w) = \int_{-\infty}^{\infty} f(\{-w \geq \alpha\})d\alpha = \int_{-\infty}^{\infty} f(\{w \leq -\alpha\})d\alpha \tag{16.78}$$

$$= \int_{-\infty}^{\infty} f(\{w \leq \alpha\})d\alpha = \int_{-\infty}^{\infty} f(\{w > \alpha\})d\alpha \tag{16.79}$$

$$= \int_{-\infty}^{\infty} f(\{w \geq \alpha\})d\alpha = \tilde{f}(w) \tag{16.80}$$

the above follows since $\int_{-\infty}^{\infty} f(\alpha)d\alpha = \int_{-\infty}^{\infty} f(a\alpha + b)d\alpha$ for any b and $a \in \pm 1$, and also since $f(A) = f(E \setminus A)$, so $f(\{w \leq \alpha\}) = f(\{w > \alpha\})$.

Lovász extension, expected value of random variable

- Recall, for $w \in \mathbb{R}^+$, we have $\tilde{f}(w) = \int_0^\infty f(\{w \geq \alpha\})d\alpha$
- Since $f(\{w \geq \alpha\}) = 0$ for $\alpha > w_1$, we have for $w \in \mathbb{R}^+$, we have $\tilde{f}(w) = \int_0^{w_1} f(\{w \geq \alpha\})d\alpha$
- For $w \in [0,1]^m$, then $\tilde{f}(w) = \int_0^1 f(\{w \geq \alpha\})d\alpha$ since $f(\{w \geq \alpha\}) = 0$ for $1 \geq \alpha > w_1$.
- Consider α as a uniform random variable on $[0,1]$ and let $h(\alpha)$ be a funciton of α. Then the expected value $\mathbb{E}[f(\alpha)] = \int_0^1 h(\alpha)d\alpha$.
- Hence, for $w \in [0,1]^m$, we can also define the Lovász extension as

$$\tilde{f}(w) = \mathbb{E}[f(\{w \geq \alpha\})] = \mathbb{E}[f(e \in E : w(e_i) \geq \alpha)] \tag{16.81}$$

where α is uniform random variable in $[0,1]$.
- This is very useful for showing results for various randomized rounding schemes when solving submodular optimization problems subject to constraints via relaxations to convex optimization problems subject to linear constraints.

Lovász extension, and polynomial time SFM

- For a long time, it was not known if general purpose submodular function minimization was doable in polynomial time.
- This was answered in the early 1980s via the help of the Lovász extension.
- The convexity of the Lovász extension, the ease of minimizing convex functions, and the fact that we can recover f from $\tilde{f}$ via $f(A) = \tilde{f}(\mathbf{1}_A)$ corresponds to why SFM is possible in polynomial time (which was first shown by Grötschel, Lovász, and Schrijver in 1988 as part of their Ellipsoid method.

Minimizing $\tilde{f}$ vs. minimizing f

In fact, we have:

Theorem 16.6.2

Let f be submodular and $\tilde{f}$ be its Lovász extension. Then $\min\{f(A) | A \subseteq E\} = \min_{w \in \{0,1\}^E} \tilde{f}(w) = \min_{w \in [0,1]^E} \tilde{f}(w)$.

Proof.

- First, since $\tilde{f}(\mathbf{1}_A) = f(A), \forall A \subseteq V$, we clearly have $\min\{f(A) | A \subseteq V\} = \min_{w \in \{0,1\}^E} \tilde{f}(w) \geq \min_{w \in [0,1]^E} \tilde{f}(w)$.
- Next, consider any $w \in [0,1]^E$, sort elements $E = \{e_1, \ldots, e_m\}$ as $w(e_1) \geq w(e_2) \geq \cdots \geq w(e_m)$, define $E_i = \{e_1, \ldots, e_i\}$, and define $\lambda_m = w(e_m)$ and $\lambda_i = w(e_i) - w(e_{i+1})$ for $i \in \{1, \ldots, m-1\}$.
- Then, as we have seen, $w = \sum_i \lambda_i \mathbf{1}_{E_i}$ and $\lambda_i \geq 0$.
- Also, $\sum_i \lambda_i = w(e_1) \leq 1$.

...

Minimizing $\tilde{f}$ vs. minimizing f

. . . cont. proof of Thm. 16.6.2.

- Note that since $f(\emptyset) = 0$, $\min\{f(A)|A \subseteq E\} \leq 0$.
- Then we have

$$\tilde{f}(w) = \int_0^1 f(\{w \geq \alpha\})d\alpha = \sum_{i=1}^m \lambda_i f(E_i) \tag{16.82}$$

$$\geq \sum_{i=1}^m \lambda_i \min_{A \subseteq E} f(A) \tag{16.83}$$

$$\geq \min_{A \subseteq E} f(A) \tag{16.84}$$

- Thus, $\min\{f(A)|A \subseteq E\} = \min_{w \in [0,1]^E} \tilde{f}(w)$.

□

Other minimizers based on min of $\tilde{f}$

- Let $w^* \in \operatorname{argmin}\left\{\tilde{f}(w)|w \in [0,1]^E\right\}$ and let $A^* \in \operatorname{argmin}\{f(A)|A \subseteq V\}$.
- Previous theorem states that $\tilde{f}(w^*) = f(A^*)$.
- Let λ_i^* be the function weights and E_i^* be the sets associated with w^*. From previous theorem, we have

$$\tilde{f}(w^*) = \sum_i \lambda_i^* f(E_i^*) = f(A^*) = \min\{f(A)|A \subseteq E\} \tag{16.85}$$

and that $f(A^*) \leq f(E_i^*), \forall i$, and that $f(A^*) \leq 0$, and $\sum_i \lambda_i \leq 1$.
- Thus, since $w^* \in [0,1]^E$, each $0 \leq \lambda_i^* \leq 1$, we have for all i such that $\lambda_i^* > 0$,

$$f(E_i^*) = f(A^*) \tag{16.86}$$

meaning such E_i^* are also minimizers of f, and $\sum_i \lambda_i = 1$.
- Hence w^* is in convex hull of incidence vectors of minimizers of f.

A bit more on level sets being minimizers

- f is normalized $f(\emptyset) = 0$, so minimizer is ≤ 0.
- We know that $f(E_i^*) \geq f(A^*)$ for all i, and $f(A^*) = \sum_i \lambda_i f(E_i^*)$.
- If $f(A^*) = 0$, then we must have $f(E_i^*) = 0$ for any i such that $\lambda_i > 0$. Otherwise, assume $f(A^*) < 0$.
- Suppose there exists an i such that $f(E_i^*) > f(A^*)$.
- Then we have

$$f(A^*) = \sum_i \lambda_i f(E_i^*) > \sum_i \lambda_i f(A^*) = f(A^*) \sum_i \lambda_i \tag{16.87}$$

  and since $f(A^*) < 0$, this means that $\sum_i \lambda_i > 1$ which is a contradiction.
- Hence, must have $f(E_i^*) = f(A^*)$ for all i.
- Hence, $\sum_i \lambda_i = 1$ since $f(A^*) = \sum_i \lambda_i f(A^*)$.