Submodular Functions, Optimization, and Applications to Machine Learning

— Spring Quarter, Lecture 17 —

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

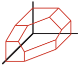
Prof. Jeff Bilmes

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

May 28th, 2014

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

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



Cumulative Outstanding Reading

- P. Wolfe - 1976
- 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/`)
- Fujishige & Isotani - 2009

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: Lovasz extension, Choquet Integration, more properties/examples of Lovasz extension, convex minimization and SFM.
- L18:
- L19:
- L20:

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

Min-Norm Point and SFM

Theorem 17.2.1

Let y^*, A_-, and A_0 be as given. Then y^* is a maximizer of the l.h.s. of Eqn. **(??)**. *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

$$(x^* + \alpha \mathbf{1}_e - \alpha \mathbf{1}_{e'})(E) = x^*(E) + \alpha - \alpha = f(E) \tag{17.1}$$

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

...

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 17.2.1

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

$$A = A_- \cup \bigcup_{a \in A_m} \mathrm{dep}(x^*, a) \tag{17.7}$$

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

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

$$\tilde{f}(w) = \max(wx : x \in P_f) \tag{17.11}$$

$$= \sum_{i=1}^{m} w(e_i) f(e_i | E_{i-1}) \tag{17.12}$$

$$= \sum_{i=1}^{m} w(e_i)(f(E_i) - f(E_{i-1})) \tag{17.13}$$

$$= w(e_m) f(E_m) + \sum_{i=1}^{m-1} (w(e_i) - w(e_{i+1})) f(E_i) \tag{17.14}$$

- 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 submodular f

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

$$\tilde{f}(w) = \max(wx : x \in P_f) \tag{17.11}$$

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

$$\tilde{f}(w) = w(e_m)f(E_m) + \sum_{i=1}^{m-1}(w(e_i) - w(e_{i+1}))f(E_i) \tag{17.12}$$

$$= \sum_{i=1}^{m} \lambda_i f(E_i) \tag{17.13}$$

  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 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

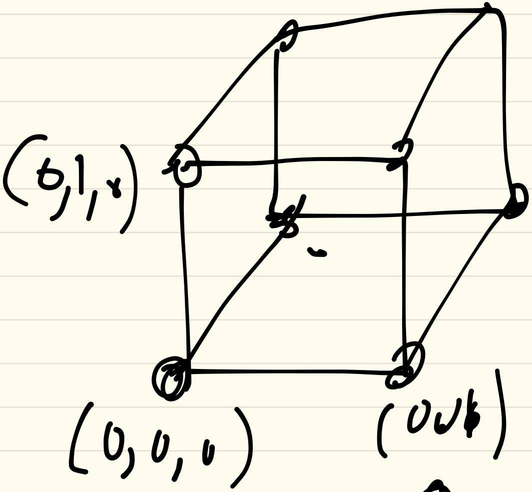
$$\tilde{f}(w) = \sum_{i=1}^{m} \lambda_i f(E_i) \tag{17.20}$$

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

$$\text{for } i \in \{1, \ldots, m\}, \quad \lambda_i = \begin{cases} w(e_i) - w(e_{i+1}) & \text{if } i < m \\ w(e_m) & \text{if } i = m \end{cases} \tag{17.21}$$

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

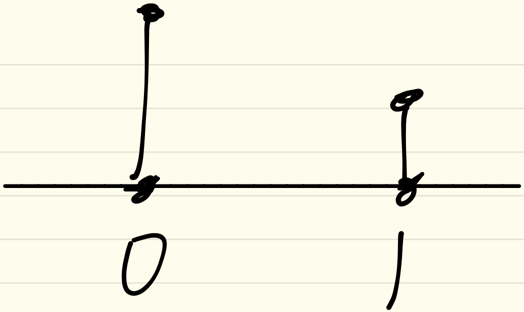
- $w = \sum_{i=1}^{m} \lambda_i \mathbf{1}_{E_i}$ is an interpolation of certain hypercube vertices.
- $\tilde{f}(w) = \sum_{i=1}^{m} \lambda_i f(E_i)$ is the associated interpolation of the values of f at sets corresponding to each hypercube vertex.

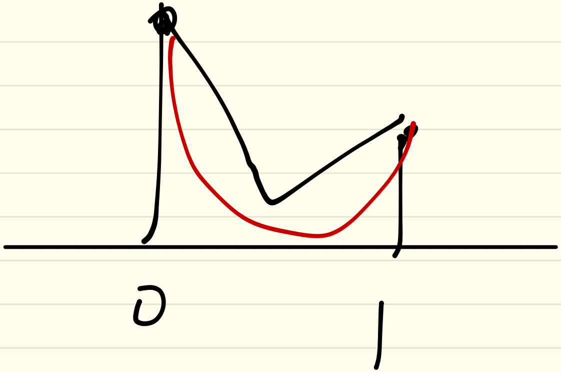


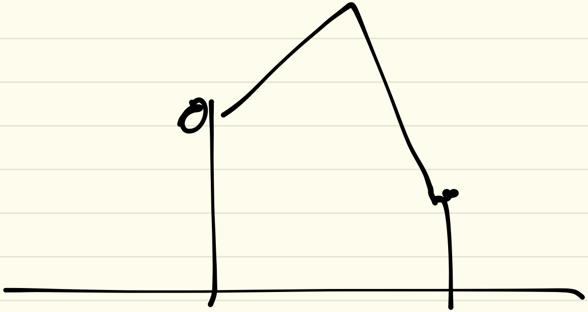
(e_1, e_2, e_3)

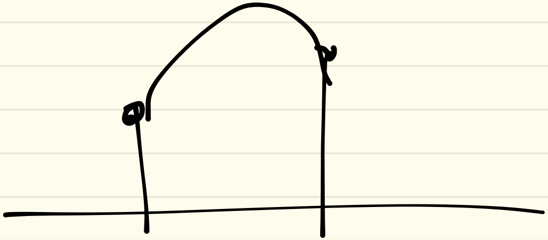
$1_A \qquad A = \{e_3\}$

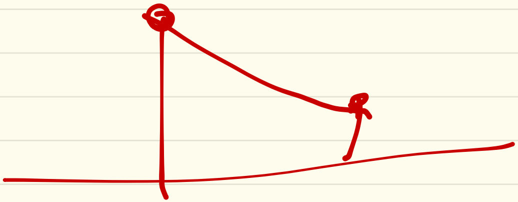
$w \in [0,1]^E$


0
1


0
1







Summary: comparison of the two extension forms

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

$$\tilde{f}(w) = \max(wx : x \in P_f) = \sum_{i=1}^{m} \lambda_i f(E_i) \tag{17.1}$$

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

Summary: comparison of the two extension forms

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

$$\tilde{f}(w) = \max(wx : x \in P_f) = \sum_{i=1}^{m} \lambda_i f(E_i) \tag{17.1}$$

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

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

$$\tilde{f}(w) = \sum_{i=1}^{m} \lambda_i f(E_i) \tag{17.2}$$

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

Summary: comparison of the two extension forms

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

$$\tilde{f}(w) = \max(wx : x \in P_f) = \sum_{i=1}^{m} \lambda_i f(E_i) \tag{17.1}$$

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

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

$$\tilde{f}(w) = \sum_{i=1}^{m} \lambda_i f(E_i) \tag{17.2}$$

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

- In both Eq. (17.1) and Eq. (17.2), we have $\tilde{f}(\mathbf{1}_A) = f(A)$, $\forall A$, but Eq. (17.2), might not be convex.

Summary: comparison of the two extension forms

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

$$\tilde{f}(w) = \max(wx : x \in P_f) = \sum_{i=1}^{m} \lambda_i f(E_i) \tag{17.1}$$

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

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

$$\tilde{f}(w) = \sum_{i=1}^{m} \lambda_i f(E_i) \tag{17.2}$$

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

- In both Eq. (17.1) and Eq. (17.2), we have $\tilde{f}(\mathbf{1}_A) = f(A)$, $\forall A$, but Eq. (17.2), might not be convex.
- Submodularity is sufficient for convexity of but is it necessary?

Lovász Extension, Submodularity and Convexity

Theorem 17.2.1

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.(**??**) 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.

...

Integration and Aggregation

- Integration is just summation (e.g., the $\int$ symbol has as its origins a sum).

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

$$\int_X f du = \sup I_X(s) \tag{17.3}$$

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.

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

$$\text{WAVG}(x) = \sum_{e \in E} x(e)w(e) \tag{17.4}$$

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

$$\text{WAVG}(x) = \sum_{e \in E} x(e) w(e) \tag{17.4}$$

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

$$\text{WAVG}(\mathbf{1}_e) = w(e) \tag{17.5}$$

Integration, Aggregation, and Weighted Averages

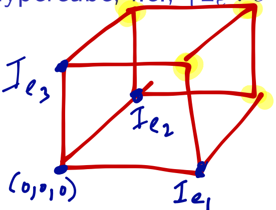
- In finite discrete spaces, Lebesgue integration is just a weighted average, and can be seen as an aggregation function.
- I.e., given a weight vector $w \in [0,1]^E$ for some finite ground set E, then for any $x \in \mathbb{R}^E$ we have

$$\text{WAVG}(x) = \sum_{e \in E} x(e)w(e) \tag{17.4}$$

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

$$\text{WAVG}(\mathbf{1}_e) = w(e) \tag{17.5}$$

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



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

$$\text{WAVG}(x) = \sum_{e \in E} x(e)w(e) \tag{17.4}$$

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

$$\text{WAVG}(\mathbf{1}_e) = w(e) \tag{17.5}$$

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

$$\text{WAVG}(x) = \sum_{e \in E} x(e)w(e) = \sum_{e \in E} x(e)\text{WAVG}(\mathbf{1}_e) \tag{17.6}$$

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

$$\text{WAVG}(x) = \sum_{e \in E} x(e) w(e) \tag{17.4}$$

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

$$\text{WAVG}(\mathbf{1}_e) = w(e) \tag{17.5}$$

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

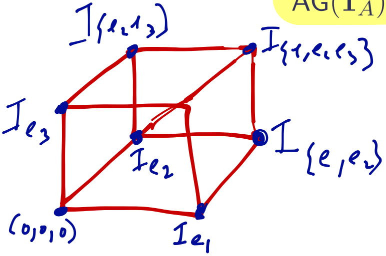
$$\text{WAVG}(x) = \sum_{e \in E} x(e) w(e) = \sum_{e \in E} x(e) \text{WAVG}(\mathbf{1}_e) \tag{17.6}$$

- Note, WAVG function is linear in the weights w, and homogeneous. $\text{WAVG}_{w_1+w_2}(x) = \text{WAVG}_{w_1}(x) + \text{WAVG}_{w_2}(x)$, $\text{WAVG}(\alpha x) = \alpha \text{WAVG}(x)$.

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 $A \subseteq E$):

$$\text{AG}(\mathbf{1}_A) = w_A \tag{17.7}$$



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 $A \subseteq E$):

$$\text{AG}(\mathbf{1}_A) = w_A \tag{17.7}$$

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

$$\text{AG}(x) = \sum_{A \subseteq E} x(A) w_A = \sum_{A \subseteq E} x(A)\text{AG}(\mathbf{1}_A) \tag{17.8}$$

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 $A \subseteq E$):

$$\text{AG}(\mathbf{1}_A) = w_A \tag{17.7}$$

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

$$\text{AG}(x) = \sum_{A \subseteq E} x(A) w_A = \sum_{A \subseteq E} x(A)\text{AG}(\mathbf{1}_A) \tag{17.8}$$

- 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

$$\text{WAVG}_{w'}(x) = \text{AG}_w(x) \tag{17.9}$$

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 $A \subseteq E$):

$$\mathsf{AG}(\mathbf{1}_A) = w_A \tag{17.7}$$

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

$$\mathsf{AG}(x) = \sum_{A \subseteq E} x(A) w_A = \sum_{A \subseteq E} x(A) \mathsf{AG}(\mathbf{1}_A) \tag{17.8}$$

- 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

$$\mathsf{WAVG}_{w'}(x) = \mathsf{AG}_w(x) \tag{17.9}$$

- 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 $A \subseteq B$.

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 $f : \{0, 1\}^m \to \mathbb{R}$.

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 $x = \mathbf{1}_A$.

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 $x \in [0,1]^m$.

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.

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 17.3.1

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

$$C_f(w) = \sum_{i=1}^{m} (w_{e_i} - w_{e_{i+1}}) f(E_i) \tag{17.10}$$

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

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

$$C_f(w) = \sum_{i=1}^{m} w(e_i)(f(E_i) - f(E_{i-1})) \tag{17.11}$$

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

Choquet integral

Definition 17.3.1

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

$$C_f(w) = \sum_{i=1}^{m} (w_{e_i} - w_{e_{i+1}}) f(E_i) \tag{17.10}$$

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

$$\sum_{k=m}^{n} a_k b_k = \sum_{k=m}^{n} A_k(b_k - b_{k+1}) + A_n b_{n+1} - A_{m-1} b_m \tag{17.12}$$

The "integral" in the Choquet integral

- Thought of as an integral over $\mathbb{R}$ of a piece-wise constant function.

The "integral" in the Choquet integral

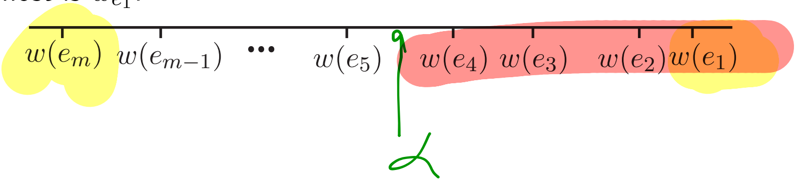
- Thought of as an integral over $\mathbb{R}$ of a piece-wise constant function.
- First note, assuming E is ordered according to descending w, so that $w(e_1) \geq w(e_2) \geq \cdots \geq w(e_{m-1}) \geq w(e_m)$, then $E_i = \{e_1, e_2, \ldots, e_i\} = \{e \in E : w_e \geq w_{e_i}\}$.

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 $w(e_1) \geq w(e_2) \geq \cdots \geq w(e_{m-1}) \geq w(e_m)$, then $E_i = \{e_1, e_2, \ldots, e_i\} = \{e \in E : w_e \geq w_{e_i}\}$.
- For any $w_{e_i} > \alpha \geq w_{e_{i+1}}$ we also have $E_i = \{e_1, e_2, \ldots, e_i\} = \{e \in E : w_e > \alpha\}$.

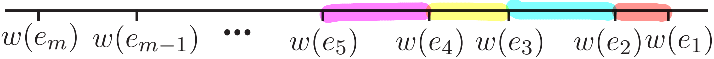
The “integral” in the Choquet integral

- 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 $w(e_1) \geq w(e_2) \geq \cdots \geq w(e_{m-1}) \geq w(e_m)$, 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}.



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 $w(e_1) \geq w(e_2) \geq \cdots \geq w(e_{m-1}) \geq w(e_m)$, then $E_i = \{e_1, e_2, \ldots, e_i\} = \{e \in E : w_e \geq w_{e_i}\}$.
- For any $w_{e_i} > \alpha \geq w_{e_{i+1}}$ we also have $E_i = \{e_1, e_2, \ldots, e_i\} = \{e \in E : w_e > \alpha\}$.
- 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

$$F_i(\alpha) = f(\{e \in E : w_e > \alpha\}) = f(E_i) \tag{17.13}$$

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:

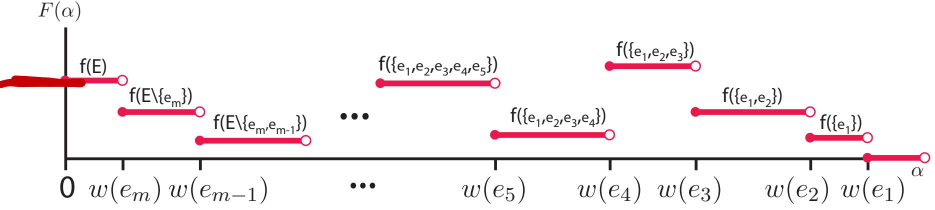
$$F(\alpha) = \begin{cases} f(E) & \text{if } \cancel{0 \le} \; \alpha < w_m \\ f(\{e \in E : w_e > \alpha\}) & \text{if } w_{e_{i+1}} \le \alpha < w_{e_i},\ i \in \{1, \dots, m-1\} \\ 0 & \text{if } w_1 < \alpha \end{cases}$$

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:

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

- 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. Why?

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 $w_{m+1} \stackrel{\text{def}}{=} 0$.

$$\tilde{f}(w) \stackrel{\text{def}}{=} \int_0^\infty F(\alpha) d\alpha \tag{17.14}$$

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 $w_{m+1} \stackrel{\text{def}}{=} 0$.

$$\tilde{f}(w) \stackrel{\text{def}}{=} \int_0^\infty F(\alpha) d\alpha \tag{17.14}$$

$$= \int_0^\infty f(\{e \in E : w_e > \alpha\}) d\alpha \tag{17.15}$$

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 $w_{m+1} \stackrel{\text{def}}{=} 0$.

$$\tilde{f}(w) \stackrel{\text{def}}{=} \int_0^\infty F(\alpha) d\alpha \tag{17.14}$$

$$= \int_0^\infty f(\{e \in E : w_e > \alpha\}) d\alpha \tag{17.15}$$

$$= \int_{w_{m+1}}^\infty f(\{e \in E : w_e > \alpha\}) d\alpha \tag{17.16}$$

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 $w_{m+1} \stackrel{\text{def}}{=} 0$.

$$\tilde{f}(w) \stackrel{\text{def}}{=} \int_0^\infty F(\alpha) d\alpha \tag{17.14}$$

$$= \int_0^\infty f(\{e \in E : w_e > \alpha\}) d\alpha \tag{17.15}$$

$$= \int_{w_{m+1}}^\infty f(\{e \in E : w_e > \alpha\}) d\alpha \tag{17.16}$$

$$= \sum_{i=1}^m \int_{w_{i+1}}^{w_i} f(\{e \in E : w_e > \alpha\}) d\alpha \tag{17.17}$$

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 $w_{m+1} \stackrel{\text{def}}{=} 0$.

$$\tilde{f}(w) \stackrel{\text{def}}{=} \int_0^\infty F(\alpha) d\alpha \tag{17.14}$$

$$= \int_0^\infty f(\{e \in E : w_e > \alpha\}) d\alpha \tag{17.15}$$

$$= \int_{w_{m+1}}^\infty f(\{e \in E : w_e > \alpha\}) d\alpha \tag{17.16}$$

$$= \sum_{i=1}^m \int_{w_{i+1}}^{w_i} f(\{e \in E : w_e > \alpha\}) d\alpha \tag{17.17}$$

$$= \sum_{i=1}^m \int_{w_{i+1}}^{w_i} f(E_i) d\alpha = \sum_{i=1}^m f(E_i)(w_i - w_{i+1}) \tag{17.18}$$

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.

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 17.3.2

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

$$\tilde{f}(w) \stackrel{\text{def}}{=} \int_0^\infty F(\alpha) d\alpha \tag{17.19}$$

where the function F is defined as before.

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 17.3.2

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

$$\tilde{f}(w) \stackrel{\text{def}}{=} \int_0^\infty F(\alpha) d\alpha \tag{17.19}$$

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:

$$\mathsf{AG}(x) = \sum_{A \subseteq E} x(A) w_A = \sum_{A \subseteq E} x(A) \mathsf{AG}(\mathbf{1}_A) \tag{17.20}$$

how does this correspond to Lovász extension?

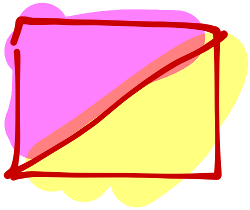
Choquet integral and aggregation

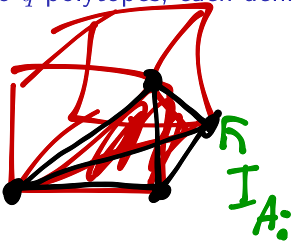
- Recall, we want to produce some notion of generalized aggregation function having the flavor of:

$$\mathsf{AG}(x) = \sum_{A \subseteq E} x(A) w_A = \sum_{A \subseteq E} x(A) \mathsf{AG}(\mathbf{1}_A) \tag{17.20}$$

  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 $\mathcal{V}_1, \mathcal{V}_2, \dots, \mathcal{V}_q$.





Choquet integral and aggregation

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

$$\mathsf{AG}(x) = \sum_{A \subseteq E} x(A) w_A = \sum_{A \subseteq E} x(A) \mathsf{AG}(\mathbf{1}_A) \tag{17.20}$$

  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.

Choquet integral and aggregation

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

$$\mathsf{AG}(x) = \sum_{A \subseteq E} x(A) w_A = \sum_{A \subseteq E} x(A) \mathsf{AG}(\mathbf{1}_A) \tag{17.20}$$

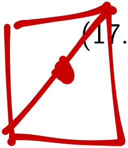
  how does this correspond to Lovász extension?

- Let us partition the hypercube $[0,1]^m$ into q polytopes, each defined by a set of vertices $\mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_q$.
- E.g., for each i, $\mathcal{V}_i = \{\mathbf{1}_{A_1}, \mathbf{1}_{A_2}, \ldots, \mathbf{1}_{A_k}\}$ (k vertices) and the convex hull of V_i defines the i^{th} polytope.
- This forms a "triangulation" of the hypercube.

Choquet integral and aggregation

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

$$\mathsf{AG}(x) = \sum_{A \subseteq E} x(A) w_A = \sum_{A \subseteq E} x(A) \mathsf{AG}(\mathbf{1}_A) \tag{17.20}$$



  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

- Most generally, for $x \in [0,1]^m$, let us define the (unique) coefficients $\alpha_0^x(A)$ and $\alpha_i^x(A)$ that define the affine transformation of the coefficients of x to be used with the particular hypercube vertex $\mathbf{1}_A$. $\in V(x)$ The affine transformation is as follows:

$$\alpha_0^x(A) + \sum_{j=1}^{m} \alpha_j^x(A) x_j \in \mathbb{R} \tag{17.21}$$

Note that many of these coefficient are often zero.

Choquet integral and aggregation

- Most generally, for $x \in [0,1]^m$, let us define the (unique) coefficients $\alpha_0^x(A)$ and $\alpha_i^x(A)$ that define the affine transformation of the coefficients of x to be used with the particular hypercube vertex $\mathbf{1}_A$. The affine transformation is as follows:

$$\alpha_0^x(A) + \sum_{j=1}^m \alpha_j^x(A) x_j \in \mathbb{R} \tag{17.21}$$

  Note that many of these coefficient are often zero.

- From this, we can define an aggregation function of the form

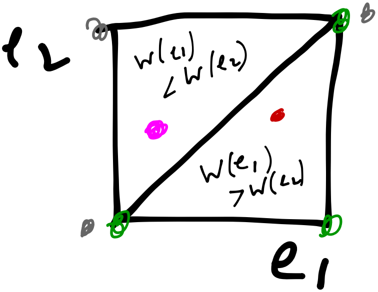
$$\text{AG}(x) \stackrel{\text{def}}{=} \sum_{A:\mathbf{1}_A \in \mathcal{V}(x)} \left(\alpha_0^x(A) + \sum_{j=1}^m \alpha_j^x(A) x_j \right) \text{AG}(\mathbf{1}_A) \tag{17.22}$$

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

$$\text{conv}(\mathcal{V}_\sigma) = \left\{ x \in [0,1]^n | x_{\sigma(1)} \geq x_{\sigma(2)} \geq \cdots \geq x_{\sigma(m)} \right\} \quad (17.23)$$

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



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

$$\mathrm{conv}(\mathcal{V}_\sigma) = \left\{x \in [0,1]^n | x_{\sigma(1)} \geq x_{\sigma(2)} \geq \cdots \geq x_{\sigma(m)}\right\} \tag{17.23}$$

  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 $\mathrm{conv}(\mathcal{V}_\sigma)$ for each permutation σ.

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

$$\operatorname{conv}(\mathcal{V}_\sigma) = \left\{x \in [0,1]^n | x_{\sigma(1)} \geq x_{\sigma(2)} \geq \cdots \geq x_{\sigma(m)}\right\} \quad (17.23)$$

  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:

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

$$\operatorname{conv}(\mathcal{V}_\sigma) = \left\{x \in [0,1]^n | x_{\sigma(1)} \geq x_{\sigma(2)} \geq \cdots \geq x_{\sigma(m)}\right\} \tag{17.23}$$

  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 17.3.3

The above linear interpolation in Eqn. (17.22) using the canonical partition yields the Lovász extension with $\alpha_0^x(A) + \sum_{j=1}^m \alpha_j^x(A) x_j = x_{\sigma_i} - x_{\sigma_{i-1}}$ *for* $A = E_i = \{e_{\sigma_1}, \ldots, e_{\sigma_i}\}$ *for appropriate order* σ*.*

$AG(I_A) = f(A)$

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

$$\mathrm{conv}(\mathcal{V}_\sigma) = \left\{x \in [0,1]^n | x_{\sigma(1)} \geq x_{\sigma(2)} \geq \cdots \geq x_{\sigma(m)}\right\} \quad (17.23)$$

  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 $\mathrm{conv}(\mathcal{V}_\sigma)$ for each permutation σ. In this case, we have:

Proposition 17.3.3

The above linear interpolation in Eqn. (17.22) using the canonical partition yields the Lovász extension with $\alpha_0^x(A) + \sum_{j=1}^m \alpha_j^x(A) x_j = x_{\sigma_i} - x_{\sigma_{i-1}}$ *for* $A = E_i = \{e_{\sigma_1}, \ldots, e_{\sigma_i}\}$ *for appropriate order* σ.

- Hence, Lovász extension is a generalized aggregation function.

Lovász extension as max over orders

- We can also write the Lovász extension as follows:

$$\tilde{f}(w) = \max_{\sigma \in \Pi_{[m]}} w^\intercal c^\sigma \tag{17.24}$$

where $\Pi_{[m]}$ is the set of $m!$ permutations of $[m] = E$, $\sigma \in \Pi_{[m]}$ is a particular permutation, and c^σ is a vector associated with permutation σ defined as:

$$c_i^\sigma = f(E_{\sigma_i}) - f(E_{\sigma_{i-1}}) \tag{17.25}$$

where $E_{\sigma_i} = \{e_{\sigma_1}, e_{\sigma_2}, \ldots, e_{\sigma_i}\}$.

Lovász extension as max over orders

- We can also write the Lovász extension as follows:

$$\tilde{f}(w) = \max_{\sigma \in \Pi_{[m]}} w^\intercal c^\sigma \tag{17.24}$$

where $\Pi_{[m]}$ is the set of $m!$ permutations of $[m] = E$, $\sigma \in \Pi_{[m]}$ is a particular permutation, and c^σ is a vector associated with permutation σ defined as:

$$c_i^\sigma = f(E_{\sigma_i}) - f(E_{\sigma_{i-1}}) \tag{17.25}$$

where $E_{\sigma_i} = \{e_{\sigma_1}, e_{\sigma_2}, \ldots, e_{\sigma_i}\}$.

- Note this immediately follows from the definition of the Lovász extension in the form:

$$\tilde{f}(w) = \max_{x \in P_f} w^\intercal x = \max_{x \in B_f} w^\intercal x \tag{17.26}$$

since we know that the maximum is achieved by an extreme point of the base B_f and all extreme points are obtained by a permutation-of-E-parameterized greedy instance.

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.

$<$

$=$

$>$

$\geq$

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).

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 $w(e_1) \geq w(e_2) \geq \cdots \geq w(e_m)$.

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 $w(e_1) \geq w(e_2) \geq \cdots \geq w(e_m)$. Also, w.l.o.g., number elements of w so that $w_1 \geq w_2 \geq \cdots \geq w_m$.

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 $w(e_1) \geq w(e_2) \geq \cdots \geq w(e_m)$. Also, w.l.o.g., number elements of w so that $w_1 \geq w_2 \geq \cdots \geq w_m$.
- We have already seen how we can define the Lovász extension for any (not necessarily submodular) function f in the following equivalent ways:

$$\tilde{f}(w) = \sum_{i=1}^{m} w(e_i) f(e_i | E_{i-1}) \tag{17.27}$$

$$= \sum_{i=1}^{m-1} f(E_i)(w(e_i) - w(e_{i+1})) + f(E)w(e_m)a \tag{17.28}$$

$$= \sum_{i=1}^{m-1} \lambda_i f(E_i) \tag{17.29}$$

Lovász extension, as integral

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

$$\tilde{f}(w) = \sum_{i=1}^{m} w(e_i) f(e_i|E_{i-1}) = \sum_{i=1}^{m} \lambda_i f(E_i) \tag{17.30}$$

$$= \sum_{i=1}^{m-1} f(E_i)(w(e_i) - w(e_{i+1})) + f(E) w(e_m) \tag{17.31}$$

$$= \int_{\min\{w_1,\dots,w_m\}}^{+\infty} f(\{w \geq \alpha\}) d\alpha + f(E) \min\{w_1, \dots, w_m\} \tag{17.32}$$

$$= \int_{0}^{+\infty} f(\{w \geq \alpha\}) d\alpha + \int_{-\infty}^{0} [f(\{w \geq \alpha\}) - f(E)] d\alpha \tag{17.33}$$

general Lovász extension, as simple integral

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

$$\tilde{f}(w) = \int_{-\infty}^{+\infty} \hat{f}(\alpha) d\alpha \tag{17.34}$$

where

$$\hat{f}(\alpha) = \begin{cases} f(\{w \geq \alpha\}) & \text{if } \alpha >= 0 \\ f(\{w \geq \alpha\}) - f(E) & \text{if } \alpha < 0 \end{cases} \tag{17.35}$$

general Lovász extension, as simple integral

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

$$\tilde{f}(w) = \int_{-\infty}^{+\infty} \hat{f}(\alpha) d\alpha \tag{17.34}$$

where

$$\hat{f}(\alpha) = \begin{cases} f(\{w \geq \alpha\}) & \text{if } \alpha >= 0 \\ f(\{w \geq \alpha\}) - f(E) & \text{if } \alpha < 0 \end{cases} \tag{17.35}$$

- So we can write it as a simple integral over the right function.

general Lovász extension, as simple integral

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

$$\tilde{f}(w) = \int_{-\infty}^{+\infty} \hat{f}(\alpha) d\alpha \tag{17.34}$$

where

$$\hat{f}(\alpha) = \begin{cases} f(\{w \geq \alpha\}) & \text{if } \alpha >= 0 \\ f(\{w \geq \alpha\}) - f(E) & \text{if } \alpha < 0 \end{cases} \tag{17.35}$$

- 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. (17.32), first note that the r.h.s. terms are the same since $w(e_m) = \min\{w_1, \ldots, w_m\}$.

Lovász extension, as integral

- To show Eqn. (17.32), first note that the r.h.s. terms are the same since $w(e_m) = \min\{w_1, \ldots, w_m\}$.
- Then, consider that, as a function of α, we have

$$f(\{w \geq \alpha\}) = \begin{cases} 0 & \text{if } \alpha > w(e_1) \\ f(E_k) & \text{if } \alpha \in (w(e_{k+1}), w(e_k)), k \in \{1, \ldots, m-1\} \\ f(E) & \text{if } \alpha < w(e_m) \end{cases} \tag{17.36}$$

we use open intervals since sets of zero measure don't change integration.

Lovász extension, as integral

- To show Eqn. (17.32), first note that the r.h.s. terms are the same since $w(e_m) = \min \{w_1, \ldots, w_m\}$.
- Then, consider that, as a function of α, we have

$$f(\{w \geq \alpha\}) = \begin{cases} 0 & \text{if } \alpha > w(e_1) \\ f(E_k) & \text{if } \alpha \in (w(e_{k+1}), w(e_k)), k \in \{1, \ldots, m-1\} \\ f(E) & \text{if } \alpha < w(e_m) \end{cases} \tag{17.36}$$

  we use open intervals since sets of zero measure don't change integration.
- Inside the integral, then, this recovers Eqn. (17.31).

Lovász extension, as integral

- To show Eqn. (17.33), start with Eqn. (17.32), note $w_m = \min\{w_1, \dots, w_m\}$, take any $\beta \leq \min\{0, w_1, \dots, w_m\}$, and form:

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

Lovász extension, as integral

- To show Eqn. (17.33), start with Eqn. (17.32), note $w_m = \min\{w_1, \ldots, w_m\}$, take any $\beta \leq \min\{0, w_1, \ldots, w_m\}$, and form:

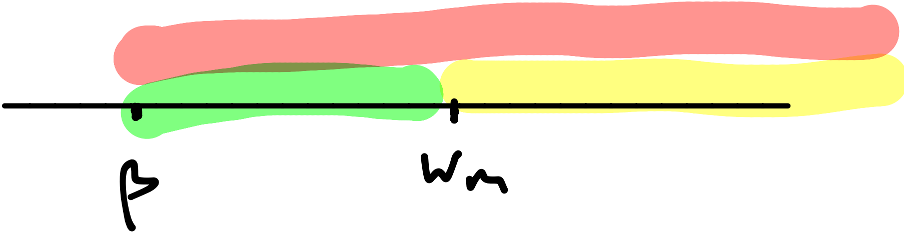
$$\tilde{f}(w) = \int_{w_m}^{+\infty} f(\{w \geq \alpha\})d\alpha + f(E)\min\{w_1, \ldots, w_m\}$$

Lovász extension, as integral

- To show Eqn. (17.33), start with Eqn. (17.32), note $w_m = \min\{w_1, \ldots, w_m\}$, take any $\beta \leq \min\{0, w_1, \ldots, w_m\}$, and form:

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

$$= \int_{\beta}^{+\infty} f(\{w \geq \alpha\})d\alpha - \int_{\beta}^{w_m} f(\{w \geq \alpha\})d\alpha + f(E)\int_0^{w_m} d\alpha$$



Lovász extension, as integral

- To show Eqn. (17.33), start with Eqn. (17.32), note $w_m = \min\{w_1, \ldots, w_m\}$, take any $\beta \leq \min\{0, w_1, \ldots, w_m\}$, and form:

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

$$= \int_{\beta}^{+\infty} f(\{w \geq \alpha\})d\alpha - \int_{\beta}^{w_m} f(\{w \geq \alpha\})d\alpha + f(E)\int_0^{w_m} d\alpha$$

$$= \int_{\beta}^{+\infty} f(\{w \geq \alpha\})d\alpha - \int_{\beta}^{w_m} f(E)d\alpha + \int_0^{w_m} f(E)d\alpha$$

Lovász extension, as integral

- To show Eqn. (17.33), start with Eqn. (17.32), note $w_m = \min\{w_1, \dots, w_m\}$, take any $\beta \leq \min\{0, w_1, \dots, w_m\}$, and form:

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

Lovász extension, as integral

- To show Eqn. (17.33), start with Eqn. (17.32), note $w_m = \min\{w_1, \dots, w_m\}$, take any $\beta \leq \min\{0, w_1, \dots, w_m\}$, and form:

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

Lovász extension, as integral

- To show Eqn. (17.33), start with Eqn. (17.32), note $w_m = \min\{w_1, \dots, w_m\}$, take any $\beta \leq \min\{0, w_1, \dots, w_m\}$, and form:

$$\begin{aligned}
\tilde{f}(w) &= \int_{w_m}^{+\infty} f(\{w \geq \alpha\})d\alpha + f(E)\min\{w_1, \dots, w_m\} \\
&= \int_{\beta}^{+\infty} f(\{w \geq \alpha\})d\alpha - \int_{\beta}^{w_m} f(\{w \geq \alpha\})d\alpha + f(E)\int_0^{w_m} d\alpha \\
&= \int_{\beta}^{+\infty} f(\{w \geq \alpha\})d\alpha - \int_{\beta}^{w_m} f(E)d\alpha + \int_0^{w_m} f(E)d\alpha \\
&= \int_{0}^{+\infty} f(\{w \geq \alpha\})d\alpha + \int_{\beta}^{0} f(\{w \geq \alpha\})d\alpha - \int_{\beta}^{0} f(E)d\alpha \\
&= \int_{0}^{+\infty} f(\{w \geq \alpha\})d\alpha + \int_{\beta}^{0} [f(\{w \geq \alpha\}) - f(E)]d\alpha \qquad (17.33)
\end{aligned}$$

and then let $\beta \to \infty$ and we get Eqn. (17.33), i.e.:

Lovász extension, as integral

- To show Eqn. (17.33), start with Eqn. (17.32), note $w_m = \min\{w_1, \dots, w_m\}$, take any $\beta \le \min\{0, w_1, \dots, w_m\}$, and form:

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

and then let $\beta \to \infty$ and we get Eqn. (17.33), i.e.:

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

Lovász extension properties

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

Lovász extension properties

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

Theorem 17.4.1

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

Lovász extension properties

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

Theorem 17.4.1

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 $\lambda \in \mathbb{R}$.*

Lovász extension properties

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

Theorem 17.4.1

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 $\tilde{f}(w) = \int_0^{+\infty} f(\{w \geq \alpha\})d\alpha$.*

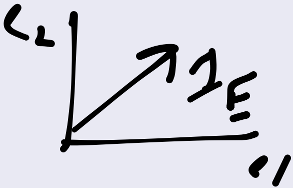
Lovász extension properties

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

Theorem 17.4.1

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 $\tilde{f}(w + \alpha \mathbf{1}_E) = \tilde{f}(w) + \alpha f(E)$.*



Lovász extension properties

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

Theorem 17.4.1

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 $\alpha \geq 0$.*

Lovász extension properties

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

Theorem 17.4.1

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 $A \subseteq E$, $\tilde{f}(\mathbf{1}_A) = f(A)$.*

Lovász extension properties

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

Theorem 17.4.1

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).*

Lovász extension properties

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

Theorem 17.4.1

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

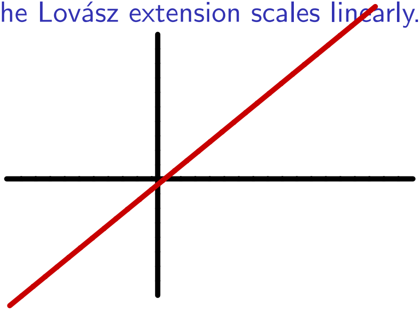
1. *Superposition of LE operator: Given f and g with Lovász extensions $\tilde{f}$ and $\tilde{g}$ then $\tilde{f} + \tilde{g}$ is the Lovász extension of $f + g$ and $\lambda \tilde{f}$ is the Lovász extension of λf for $\lambda \in \mathbb{R}$.*
2. *If $w \in \mathbb{R}_+^E$ then $\tilde{f}(w) = \int_0^{+\infty} f(\{w \geq \alpha\}) d\alpha$.*
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 $E^1 \cup E^2 \cup \cdots \cup E^k$ of E and $w = \sum_{i=1}^k \gamma_i \mathbf{1}_{E_k}$ with $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_k$, and with $E^{1:i} = E^1 \cup E^2 \cup \cdots \cup E^i$, 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 $\tilde{f}(w + \alpha \mathbf{1}_E) = \tilde{f}(w) + \alpha f(E)$.

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.



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. (17.30) through (17.33), most of the above properties are relatively easy to derive.

Lovász extension properties

- Given Eqns. (17.30) through (17.33), 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

$$\tilde{f}(-w) \tag{17.39}$$

Lovász extension properties

- Given Eqns. (17.30) through (17.33), 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

$$\tilde{f}(-w) = \int_{-\infty}^{\infty} f(\{-w \geq \alpha\})d\alpha \tag{17.39}$$

Lovász extension properties

- Given Eqns. (17.30) through (17.33), 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

$$\tilde{f}(-w) = \int_{-\infty}^{\infty} f(\{-w \geq \alpha\})d\alpha = \int_{-\infty}^{\infty} f(\{w \leq -\alpha\})d\alpha \quad (17.37)$$

(17.39)

Lovász extension properties

- Given Eqns. (17.30) through (17.33), 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

$$\tilde{f}(-w) = \int_{-\infty}^{\infty} f(\{-w \geq \alpha\}) d\alpha = \int_{-\infty}^{\infty} f(\{w \leq -\alpha\}) d\alpha \quad (17.37)$$

$$\stackrel{(a)}{=} \int_{-\infty}^{\infty} f(\{w \leq \alpha\}) d\alpha$$

(17.39)

Lovász extension properties

- Given Eqns. (17.30) through (17.33), 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

$$\tilde{f}(-w) = \int_{-\infty}^{\infty} f(\{-w \geq \alpha\}) d\alpha = \int_{-\infty}^{\infty} f(\{w \leq -\alpha\}) d\alpha \tag{17.37}$$

$$\stackrel{(a)}{=} \int_{-\infty}^{\infty} f(\{w \leq \alpha\}) d\alpha \stackrel{(b)}{=} \int_{-\infty}^{\infty} f(\{w > \alpha\}) d\alpha \tag{17.38}$$

(17.39)

Lovász extension properties

- Given Eqns. (17.30) through (17.33), 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

$$\tilde{f}(-w) = \int_{-\infty}^{\infty} f(\{-w \geq \alpha\})d\alpha = \int_{-\infty}^{\infty} f(\{w \leq -\alpha\})d\alpha \tag{17.37}$$

$$\stackrel{(a)}{=} \int_{-\infty}^{\infty} f(\{w \leq \alpha\})d\alpha \stackrel{(b)}{=} \int_{-\infty}^{\infty} f(\{w > \alpha\})d\alpha \tag{17.38}$$

$$= \int_{-\infty}^{\infty} f(\{w \geq \alpha\})d\alpha \tag{17.39}$$

Lovász extension properties

- Given Eqns. (17.30) through (17.33), 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

$$\tilde{f}(-w) = \int_{-\infty}^{\infty} f(\{-w \geq \alpha\}) d\alpha = \int_{-\infty}^{\infty} f(\{w \leq -\alpha\}) d\alpha \tag{17.37}$$

$$\stackrel{(a)}{=} \int_{-\infty}^{\infty} f(\{w \leq \alpha\}) d\alpha \stackrel{(b)}{=} \int_{-\infty}^{\infty} f(\{w > \alpha\}) d\alpha \tag{17.38}$$

$$= \int_{-\infty}^{\infty} f(\{w \geq \alpha\}) d\alpha = \tilde{f}(w) \tag{17.39}$$

Lovász extension properties

- Given Eqns. (17.30) through (17.33), 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

$$\tilde{f}(-w) = \int_{-\infty}^{\infty} f(\{-w \geq \alpha\})d\alpha = \int_{-\infty}^{\infty} f(\{w \leq -\alpha\})d\alpha \quad (17.37)$$

$$\stackrel{(a)}{=} \int_{-\infty}^{\infty} f(\{w \leq \alpha\})d\alpha \stackrel{(b)}{=} \int_{-\infty}^{\infty} f(\{w > \alpha\})d\alpha \quad (17.38)$$

$$= \int_{-\infty}^{\infty} f(\{w \geq \alpha\})d\alpha = \tilde{f}(w) \quad (17.39)$$

Equality (a) 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 equality (b) follows since $f(A) = f(E \setminus A)$, so $f(\{w \leq \alpha\}) = f(\{w > \alpha\})$.

Lovász extension, expected value of random variable

- Recall, for $w \in \mathbb{R}_+^E$, we have $\tilde{f}(w) = \int_0^\infty f(\{w \geq \alpha\})d\alpha$

Lovász extension, expected value of random variable

- Recall, for $w \in \mathbb{R}_+^E$, we have $\tilde{f}(w) = \int_0^\infty f(\{w \geq \alpha\}) d\alpha$
- Since $f(\{w \geq \alpha\}) = 0$ for $\alpha > w_1 \geq w_\ell$, we have for $w \in \mathbb{R}_+^E$, we have $\tilde{f}(w) = \int_0^{w_1} f(\{w \geq \alpha\}) d\alpha$

Lovász extension, expected value of random variable

- Recall, for $w \in \mathbb{R}_+^E$, we have $\tilde{f}(w) = \int_0^\infty f(\{w \geq \alpha\})d\alpha$
- Since $f(\{w \geq \alpha\}) = 0$ for $\alpha > w_1 \geq w_\ell$, we have for $w \in \mathbb{R}_+^E$, we have $\tilde{f}(w) = \int_0^{w_1} f(\{w \geq \alpha\})d\alpha$
- For $w \in [0,1]^E$, then $\tilde{f}(w) = \int_0^{w_1} f(\{w \geq \alpha\})d\alpha = \int_0^1 f(\{w \geq \alpha\})d\alpha$ since $f(\{w \geq \alpha\}) = 0$ for $1 \geq \alpha > w_1$.

Lovász extension, expected value of random variable

- Recall, for $w \in \mathbb{R}_+^E$, we have $\tilde{f}(w) = \int_0^\infty f(\{w \geq \alpha\}) d\alpha$
- Since $f(\{w \geq \alpha\}) = 0$ for $\alpha > w_1 \geq w_\ell$, we have for $w \in \mathbb{R}_+^E$, we have $\tilde{f}(w) = \int_0^{w_1} f(\{w \geq \alpha\}) d\alpha$
- For $w \in [0,1]^E$, then $\tilde{f}(w) = \int_0^{w_1} f(\{w \geq \alpha\}) d\alpha = \int_0^1 f(\{w \geq \alpha\}) d\alpha$ 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 $\mathbb{E}[f(\alpha)] = \int_0^1 h(\alpha) d\alpha$.

Lovász extension, expected value of random variable

- Recall, for $w \in \mathbb{R}_+^E$, we have $\tilde{f}(w) = \int_0^\infty f(\{w \geq \alpha\}) d\alpha$
- Since $f(\{w \geq \alpha\}) = 0$ for $\alpha > w_1 \geq w_\ell$, we have for $w \in \mathbb{R}_+^E$, we have $\tilde{f}(w) = \int_0^{w_1} f(\{w \geq \alpha\}) d\alpha$
- For $w \in [0,1]^E$, then $\tilde{f}(w) = \int_0^{w_1} f(\{w \geq \alpha\}) d\alpha = \int_0^1 f(\{w \geq \alpha\}) d\alpha$ 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

$$\tilde{f}(w) = \mathbb{E}[f(\{w \geq \alpha\})] = \mathbb{E}[f(e \in E : w(e_i) \geq \alpha)] \tag{17.40}$$

where α is uniform random variable in $[0,1]$.

Lovász extension, expected value of random variable

- Recall, for $w \in \mathbb{R}_+^E$, we have $\tilde{f}(w) = \int_0^\infty f(\{w \geq \alpha\})d\alpha$
- Since $f(\{w \geq \alpha\}) = 0$ for $\alpha > w_1 \geq w_\ell$, we have for $w \in \mathbb{R}_+^E$, we have $\tilde{f}(w) = \int_0^{w_1} f(\{w \geq \alpha\})d\alpha$
- For $w \in [0,1]^E$, then $\tilde{f}(w) = \int_0^{w_1} f(\{w \geq \alpha\})d\alpha = \int_0^1 f(\{w \geq \alpha\})d\alpha$ 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 $\mathbb{E}[h(\alpha)] = \int_0^1 h(\alpha)d\alpha$.
- Hence, for $w \in [0,1]^m$, we can also define the Lovász extension as

$$\tilde{f}(w) = \mathbb{E}_{p(\alpha)}[\underbrace{f(\{w \geq \alpha\})}_{h(\alpha)}] = \mathbb{E}_{p(\alpha)}[\underbrace{f(e \in E : w(e_i) \geq \alpha)}_{h(\alpha)}] \quad (17.40) \quad = \int p(\alpha)h(\alpha)d\alpha$$

  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.

Ellipsoid algorithm, and polynomial time SFM

- For a long time, it was not known if general purpose submodular function minimization was doable in polynomial time.

Ellipsoid algorithm, 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 Edmonds's greedy algorithm from 1970.

Ellipsoid algorithm, 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 Edmonds's greedy algorithm from 1970. Let $C \subseteq \mathbb{R}^V$ be a non-empty convex compact set.

Ellipsoid algorithm, 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 Edmonds's greedy algorithm from 1970. Let $C \subseteq \mathbb{R}^V$ be a non-empty convex compact set.

Definition 17.5.1 ((strong) optimization problem)

Given $c \in \mathbb{R}^V$, find a vector $x \in C$ that maximizes $c^\intercal x$ on C. I.e., solve

$$\max_{x \in C} c^\intercal x \tag{17.41}$$

Ellipsoid algorithm, 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 Edmonds's greedy algorithm from 1970. Let $C \subseteq \mathbb{R}^V$ be a non-empty convex compact set.

Definition 17.5.1 ((strong) optimization problem)

Given $c \in \mathbb{R}^V$, find a vector $x \in C$ that maximizes $c^\intercal x$ on C. I.e., solve

$$\max_{x \in C} c^\intercal x \tag{17.41}$$

Definition 17.5.2 ((strong) separation problem)

Given a vector $y \in \mathbb{R}^V$, decide if $y \in C$, and if not, find a hyperplane that separates y from C. I.e., find vector $c \in \mathbb{R}^V$ such that:

$$c^\intercal y > \max_{x \in C} c^\intercal x \tag{17.42}$$

y C

Ellipsoid algorithm, and polynomial time SFM

- We have the following important theorem:

Ellipsoid algorithm, and polynomial time SFM

- We have the following important theorem:

Theorem 17.5.3 (Grötschel, Lovász, and Schrijver, 1981)

Let $\mathcal{C}$ be set of convex sets. Then there is a polynomial-time algorithm to solve the separation problem for the members of $\mathcal{C}$ iff there is a polynomial-time algorithm to solve the optimization problem for the members of $\mathcal{C}$.

Ellipsoid algorithm, and polynomial time SFM

- We have the following important theorem:

Theorem 17.5.3 (Grötschel, Lovász, and Schrijver, 1981)

Let $\mathcal{C}$ be set of convex sets. Then there is a polynomial-time algorithm to solve the separation problem for the members of $\mathcal{C}$ iff there is a polynomial-time algorithm to solve the optimization problem for the members of $\mathcal{C}$.

- We saw already that the greedy algorithm solves the storng separation problem for polymatroidal polytopes.

Ellipsoid algorithm, and polynomial time SFM

- We have the following important theorem:

Theorem 17.5.3 (Grötschel, Lovász, and Schrijver, 1981)

Let $\mathcal{C}$ be set of convex sets. Then there is a polynomial-time algorithm to solve the separation problem for the members of $\mathcal{C}$ iff there is a polynomial-time algorithm to solve the optimization problem for the members of $\mathcal{C}$.

- We saw already that the greedy algorithm solves the storng separation problem for polymatroidal polytopes.
- The ellipsoid algorithm first bounds a polytope P with an ellipsoid, and then creates a sequence of elipsoids of exponentially decreasing volume which are used to address a P membership problem.

Ellipsoid algorithm, and polynomial time SFM

- We have the following important theorem:

Theorem 17.5.3 (Grötschel, Lovász, and Schrijver, 1981)

Let $\mathcal{C}$ be set of convex sets. Then there is a polynomial-time algorithm to solve the separation problem for the members of $\mathcal{C}$ iff there is a polynomial-time algorithm to solve the optimization problem for the members of $\mathcal{C}$.

- We saw already that the greedy algorithm solves the storng separation problem for polymatroidal polytopes.
- The ellipsoid algorithm first bounds a polytope P with an ellipsoid, and then creates a sequence of elipsoids of exponentially decreasing volume which are used to address a P membership problem.
- This is sufficient to show that we can solve SFM in polynomial time!

Ellipsoid algorithm, and polynomial time SFM

- We have the following important theorem:

Theorem 17.5.3 (Grötschel, Lovász, and Schrijver, 1981)

Let $\mathcal{C}$ be set of convex sets. Then there is a polynomial-time algorithm to solve the separation problem for the members of $\mathcal{C}$ iff there is a polynomial-time algorithm to solve the optimization problem for the members of $\mathcal{C}$.

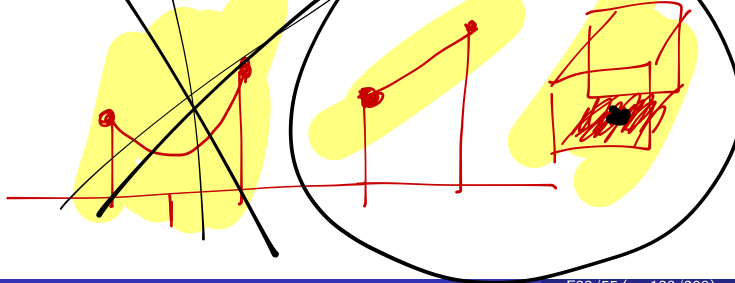
- We saw already that the greedy algorithm solves the storng separation problem for polymatroidal polytopes.
- The ellipsoid algorithm first bounds a polytope P with an ellipsoid, and then creates a sequence of elipsoids of exponentially decreasing volume which are used to address a P membership problem.
- This is sufficient to show that we can solve SFM in polynomial time!
- See also, the book: Grötschel, Lovász, and Schrijver, "Geometric Algorithms and Combinatorial Optimization"

Convex minimization and SFM

- SFM is also related to the convexity of the Lovász extension, the ease of minimizing convex functions.

Convex minimization and SFM

- SFM is also related to the convexity of the Lovász extension, the ease of minimizing convex functions.
- Also, since we can recover f from $\tilde{f}$ via $f(A) = \tilde{f}(\mathbf{1}_A)$, and (as we will see) get discrete solutions from continuous convex minimization solution.



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

In fact, we have:

Theorem 17.5.4

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

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

In fact, we have:

Theorem 17.5.4

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

...

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

In fact, we have:

Theorem 17.5.4

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 $E = \{e_1, \dots, e_m\}$ as $w(e_1) \geq w(e_2) \geq \cdots \geq w(e_m)$, define $E_i = \{e_1, \dots, e_i\}$, and define $\lambda_m = w(e_m)$ and $\lambda_i = w(e_i) - w(e_{i+1})$ for $i \in \{1, \dots, m-1\}$.

...

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

In fact, we have:

Theorem 17.5.4

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 $\lambda_i \geq 0$.

...

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

In fact, we have:

Theorem 17.5.4

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. 17.5.4.

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

□

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

...cont. proof of Thm. 17.5.4.

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

$$\tilde{f}(w) = \int_0^1 f(\{w \geq \alpha\}) d\alpha = \sum_{i=1}^{m} \lambda_i f(E_i) \tag{17.43}$$

$$\geq \sum_{i=1}^{m} \lambda_i \min_{A \subseteq E} f(A) \tag{17.44}$$

(a)

$$\geq \min_{A \subseteq E} f(A) \tag{17.45}$$

≤ 0

$\sum_i \lambda_i \leq 1$

□

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

...cont. proof of Thm. 17.5.4.

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

$$\tilde{f}(w) = \int_0^1 f(\{w \geq \alpha\}) d\alpha = \sum_{i=1}^m \lambda_i f(E_i) \tag{17.43}$$

$$\geq \sum_{i=1}^m \lambda_i \min_{A \subseteq E} f(A) \tag{17.44}$$

$$\geq \min_{A \subseteq E} f(A) \tag{17.45}$$

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

$\square$

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 $A^* \in \operatorname{argmin}\{f(A) | A \subseteq V\}$.

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 $\tilde{f}(w^*) = f(A^*)$.

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

$$\tilde{f}(w^*) = \sum_i \lambda_i^* f(E_i^*) = f(A^*) = \min\{f(A) | A \subseteq E\} \tag{17.46}$$

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

$$\tilde{f}(w^*) = \sum_i \lambda_i^* f(E_i^*) = f(A^*) = \min\{f(A) | A \subseteq E\} \tag{17.46}$$

and that $f(A^*) \leq f(E_i^*), \forall i,$

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

$$\tilde{f}(w^*) = \sum_i \lambda_i^* f(E_i^*) = f(A^*) = \min\{f(A) | A \subseteq E\} \tag{17.46}$$

and that $f(A^*) \leq f(E_i^*), \forall i$, and that $f(A^*) \leq 0$,

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

$$\tilde{f}(w^*) = \sum_i \lambda_i^* f(E_i^*) = f(A^*) = \min\{f(A) | A \subseteq E\} \tag{17.46}$$

and that $f(A^*) \leq f(E_i^*), \forall i$, and that $f(A^*) \leq 0$, and $\sum_i \lambda_i \leq 1$.

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

$$\tilde{f}(w^*) = \sum_i \lambda_i^* f(E_i^*) = f(A^*) = \min\{f(A) | A \subseteq E\} \tag{17.46}$$

  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 $\lambda_i^* > 0$,

$$f(E_i^*) = f(A^*) \tag{17.47}$$

  meaning such E_i^* are also minimizers of f, and $\sum_i \lambda_i = 1$.

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

$$\tilde{f}(w^*) = \sum_i \lambda_i^* f(E_i^*) = f(A^*) = \min\{f(A) | A \subseteq E\} \tag{17.46}$$

  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 $\lambda_i^* > 0$,

$$f(E_i^*) = f(A^*) \tag{17.47}$$

  meaning such E_i^* are also minimizers of f, and $\sum_i \lambda_i = 1$.
- Note that the negative of $f(A^*)$ is crucial here. See next slide that further explains this.

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

$$\tilde{f}(w^*) = \sum_i \lambda_i^* f(E_i^*) = f(A^*) = \min\{f(A) | A \subseteq E\} \qquad (17.46)$$

  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 $\lambda_i^* > 0$,

$$f(E_i^*) = f(A^*) \qquad (17.47)$$

  meaning such E_i^* are also minimizers of f, and $\sum_i \lambda_i = 1$.
- Note that the negative of $f(A^*)$ is crucial here. See next slide that further explains this.
- Hence $w^* = \sum_i \lambda_i^* \mathbf{1}_{E_i}$ is in convex hull of incidence vectors of

A bit more on level sets being minimizers

- f is normalized $f(\emptyset) = 0$, so minimizer is ≤ 0.

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 $f(A^*) = \sum_i \lambda_i f(E_i^*)$.

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 $\overset{0}{\cancel{f(A^*)}} = \sum_i \lambda_i f(E_i^*)$.
- If $f(A^*) = 0$, then we must have $f(E_i^*) = 0$ for any i such that $\lambda_i > 0$. Otherwise, assume $f(A^*) < 0$.

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 $f(E_i^*) > f(A^*)$.

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

$$f(A^*) = \sum_i \lambda_i f(E_i^*) > \sum_i \lambda_i f(A^*) = f(A^*) \sum_i \lambda_i \tag{17.48}$$

and since $f(A^*) < 0$, this means that $\sum_i \lambda_i > 1$ which is a contradiction.

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

$$f(A^*) = \sum_i \lambda_i f(E_i^*) > \sum_i \lambda_i f(A^*) = f(A^*) \sum_i \lambda_i \tag{17.48}$$

  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.

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

$$f(A^*) = \sum_i \lambda_i f(E_i^*) > \sum_i \lambda_i f(A^*) = f(A^*) \sum_i \lambda_i \tag{17.48}$$

  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 $f(A^*) = \sum_i \lambda_i f(A^*)$.

Alternate way to see Equation 17.47

- We know $f(A^*) \leq 0$. Consider two cases in Equation 17.47.

Alternate way to see Equation 17.47

- We know $f(A^*) \leq 0$. Consider two cases in Equation 17.47.
- Case 1: $f(A^*) = 0$. Then for any i with $\lambda_i > 0$ we must have $f(E_i) = 0$ as well for all i since $f(A^*) \leq f(E_i)$.

Alternate way to see Equation 17.47

- We know $f(A^*) \leq 0$. Consider two cases in Equation 17.47.
- Case 1: $f(A^*) = 0$. Then for any i with $\lambda_i > 0$ we must have $f(E_i) = 0$ as well for all i since $f(A^*) \leq f(E_i)$.
- Case 2 is where $f(A^*) < 0$. In this second case, we have

$$0 > f(A^*) = \sum_i \lambda_i f(E_i) \geq \sum_i \lambda_i f(A^*) \tag{17.49}$$

$$\stackrel{(a)}{\geq} \sum_i \lambda_i f(A^*) + (1 - \bar{\lambda}) f(A^*) = f(A^*) \tag{17.50}$$

where $\bar{\lambda} = \sum_i \lambda_i$ and $(1 - \bar{\lambda}) \geq 0$ and where (a) follows since $f(A^*) < 0$.

Alternate way to see Equation 17.47

- We know $f(A^*) \leq 0$. Consider two cases in Equation 17.47.
- Case 1: $f(A^*) = 0$. Then for any i with $\lambda_i > 0$ we must have $f(E_i) = 0$ as well for all i since $f(A^*) \leq f(E_i)$.
- Case 2 is where $f(A^*) < 0$. In this second case, we have

$$0 > f(A^*) = \sum_i \lambda_i f(E_i) \geq \sum_i \lambda_i f(A^*) \tag{17.49}$$

$$\stackrel{(a)}{\geq} \sum_i \lambda_i f(A^*) + (1 - \bar{\lambda}) f(A^*) = f(A^*) \tag{17.50}$$

  where $\bar{\lambda} = \sum_i \lambda_i$ and $(1 - \bar{\lambda}) \geq 0$ and where (a) follows since $f(A^*) < 0$.
- Hence, all inequalities must be equalities, which means that we must have that $\bar{\lambda} = 1$.

θ-rounding the L.E. minimum

We can also view the above as a form of rounding a continuous convex relaxation to the problem.

Definition 17.5.5 (θ-rounding)

Given vector $x \in [0,1]^E$, choose $\theta \in (0,1)$ and define a set corresponding to elements above θ, i.e.,

$$\hat{X}_\theta = \{i : \hat{x}(i) \geq \theta\} \triangleq \{\hat{x} \geq \theta\} \tag{17.51}$$

Lemma 17.5.6 (Fujishige-2005)

Given a continuous minimizer $x^* \in \operatorname{argmin}_{x \in [0,1]^n} \tilde{f}(x)$, the discrete minimizers are exactly the maximal chain of sets $\emptyset \subset X_{\theta_1} \subset \ldots X_{\theta_k}$ obtained by θ-rounding x^*, for $\theta_j \in (0,1)$.

Simple expressions for Lovász E. with $m = 2$, $E = \{1, 2\}$

- If $w_1 \geq w_2$, then

$$\tilde{f}(w) = w_1 f(\{1\}) + w_2 f(\{2\}|\{1\}) \tag{17.52}$$

$$= (w_1 - w_2) f(\{1\}) + w_2 f(\{1, 2\}) \tag{17.53}$$

Simple expressions for Lovász E. with $m = 2$, $E = \{1, 2\}$

- If $w_1 \geq w_2$, then

$$\tilde{f}(w) = w_1 f(\{1\}) + w_2 f(\{2\}|\{1\}) \tag{17.52}$$
$$= (w_1 - w_2) f(\{1\}) + w_2 f(\{1,2\}) \tag{17.53}$$

- If $w_1 \leq w_2$, then

$$\tilde{f}(w) = w_2 f(\{2\}) + w_1 f(\{1\}|\{2\}) \tag{17.54}$$
$$= (w_2 - w_1) f(\{2\}) + w_1 f(\{1,2\}) \tag{17.55}$$

Simple expressions for Lovász E. with $m = 2$, $E = \{1, 2\}$

- If $w_1 \geq w_2$, then

$$\tilde{f}(w) = w_1 f(\{1\}) + w_2 f(\{2\}|\{1\}) \tag{17.56}$$

$$= (w_1 - w_2) f(\{1\}) + w_2 f(\{1, 2\}) \tag{17.57}$$

$$= \frac{1}{2} f(1)(w_1 - w_2) + \frac{1}{2} f(1)(w_1 - w_2) \tag{17.58}$$

$$+ \frac{1}{2} f(\{1, 2\})(w_1 + w_2) - \frac{1}{2} f(\{1, 2\})(w_1 - w_2) \tag{17.59}$$

$$+ \frac{1}{2} f(2)(w_1 - w_2) + \frac{1}{2} f(2)(w_2 - w_1) \tag{17.60}$$

Simple expressions for Lovász E. with $m = 2$, $E = \{1, 2\}$

- If $w_1 \geq w_2$, then

$$\tilde{f}(w) = w_1 f(\{1\}) + w_2 f(\{2\}|\{1\}) \tag{17.56}$$

$$= (w_1 - w_2) f(\{1\}) + w_2 f(\{1, 2\}) \tag{17.57}$$

$$= \frac{1}{2} f(1)(w_1 - w_2) + \frac{1}{2} f(1)(w_1 - w_2) \tag{17.58}$$

$$+ \frac{1}{2} f(\{1, 2\})(w_1 + w_2) - \frac{1}{2} f(\{1, 2\})(w_1 - w_2) \tag{17.59}$$

$$+ \frac{1}{2} f(2)(w_1 - w_2) + \frac{1}{2} f(2)(w_2 - w_1) \tag{17.60}$$

- A similar (symmetric) expression holds when $w_1 \leq w_2$.

Simple expressions for Lovász E. with $m = 2$, $E = \{1, 2\}$

- This gives, for general w_1, w_2, that

$$\tilde{f}(w) = \frac{1}{2}\left(f(\{1\}) + f(\{2\}) - f(\{1,2\})\right)|w_1 - w_2| \tag{17.61}$$

$$+ \frac{1}{2}\left(f(\{1\}) - f(\{2\}) + f(\{1,2\})\right) w_1 \tag{17.62}$$

$$+ \frac{1}{2}\left(-f(\{1\}) + f(\{2\}) + f(\{1,2\})\right) w_2 \tag{17.63}$$

$$= -\left(f(\{1\}) + f(\{2\}) - f(\{1,2\})\right) \min\{w_1, w_2\} \tag{17.64}$$

$$+ f(\{1\})w_1 + f(\{2\})w_2 \tag{17.65}$$

Simple expressions for Lovász E. with $m = 2$, $E = \{1, 2\}$

- This gives, for general w_1, w_2, that

$$\tilde{f}(w) = \frac{1}{2}\left(f(\{1\}) + f(\{2\}) - f(\{1,2\})\right)|w_1 - w_2| \tag{17.61}$$
$$+ \frac{1}{2}\left(f(\{1\}) - f(\{2\}) + f(\{1,2\})\right) w_1 \tag{17.62}$$
$$+ \frac{1}{2}\left(-f(\{1\}) + f(\{2\}) + f(\{1,2\})\right) w_2 \tag{17.63}$$
$$= -\left(f(\{1\}) + f(\{2\}) - f(\{1,2\})\right) \min\{w_1, w_2\} \tag{17.64}$$
$$+ f(\{1\})w_1 + f(\{2\})w_2 \tag{17.65}$$

- Thus, if $f(A) = H(X_A)$ is the entropy function, we have $\tilde{f}(w) = H(e_1)w_1 + H(e_2)w_2 - I(e_1; e_2)\min\{w_1, w_2\}$ which must be convex in w, where $I(e_1; e_2)$ is the mutual information.

Simple expressions for Lovász E. with $m = 2$, $E = \{1, 2\}$

- This gives, for general w_1, w_2, that

$$\tilde{f}(w) = \frac{1}{2}\left(f(\{1\}) + f(\{2\}) - f(\{1,2\})\right)|w_1 - w_2| \tag{17.61}$$
$$+ \frac{1}{2}\left(f(\{1\}) - f(\{2\}) + f(\{1,2\})\right) w_1 \tag{17.62}$$
$$+ \frac{1}{2}\left(-f(\{1\}) + f(\{2\}) + f(\{1,2\})\right) w_2 \tag{17.63}$$
$$= -\left(f(\{1\}) + f(\{2\}) - f(\{1,2\})\right) \min\{w_1, w_2\} \tag{17.64}$$
$$+ f(\{1\})w_1 + f(\{2\})w_2 \tag{17.65}$$

- Thus, if $f(A) = H(X_A)$ is the entropy function, we have $\tilde{f}(w) = H(e_1)w_1 + H(e_2)w_2 - I(e_1; e_2) \min\{w_1, w_2\}$ which must be convex in w, where $I(e_1; e_2)$ is the mutual information.
- This "simple" but general form of the Lovász extension with $m = 2$ can be useful.

Example: $m = 2$, $E = \{1, 2\}$, contours

- If $w_1 \geq w_2$, then

$$\tilde{f}(w) = w_1 f(\{1\}) + w_2 f(\{2\}|\{1\}) \tag{17.66}$$

Example: $m = 2$, $E = \{1, 2\}$, contours

- If $w_1 \geq w_2$, then

$$\tilde{f}(w) = w_1 f(\{1\}) + w_2 f(\{2\}|\{1\}) \tag{17.66}$$

  - If $w = (1,0)/f(\{1\}) = \Big(1/f(\{1\}), 0\Big)$ then $\tilde{f}(w) = 1$.

Example: $m = 2$, $E = \{1, 2\}$, contours

- If $w_1 \geq w_2$, then

$$\tilde{f}(w) = w_1 f(\{1\}) + w_2 f(\{2\}|\{1\}) \tag{17.66}$$

  - If $w = (1, 0)/f(\{1\}) = \Big(1/f(\{1\}), 0\Big)$ then $\tilde{f}(w) = 1$.
  - If $w = (1, 1)/f(\{1, 2\})$ then $\tilde{f}(w) = 1$.

Example: $m = 2$, $E = \{1, 2\}$, contours

- If $w_1 \geq w_2$, then

$$\tilde{f}(w) = w_1 f(\{1\}) + w_2 f(\{2\}|\{1\}) \tag{17.66}$$

  - If $w = (1, 0)/f(\{1\}) = \Big(1/f(\{1\}), 0\Big)$ then $\tilde{f}(w) = 1$.
  - If $w = (1, 1)/f(\{1, 2\})$ then $\tilde{f}(w) = 1$.

- If $w_1 \leq w_2$, then

$$\tilde{f}(w) = w_2 f(\{2\}) + w_1 f(\{1\}|\{2\}) \tag{17.67}$$

Example: $m = 2$, $E = \{1, 2\}$, contours

- If $w_1 \geq w_2$, then

$$\tilde{f}(w) = w_1 f(\{1\}) + w_2 f(\{2\}|\{1\}) \tag{17.66}$$

  - If $w = (1, 0)/f(\{1\}) = \Big(1/f(\{1\}), 0\Big)$ then $\tilde{f}(w) = 1$.
  - If $w = (1, 1)/f(\{1, 2\})$ then $\tilde{f}(w) = 1$.

- If $w_1 \leq w_2$, then

$$\tilde{f}(w) = w_2 f(\{2\}) + w_1 f(\{1\}|\{2\}) \tag{17.67}$$

  - If $w = (0, 1)/f(\{2\}) = (0, 1/f(\{2\}))$ then $\tilde{f}(w) = 1$.

Example: $m = 2$, $E = \{1, 2\}$, contours

- If $w_1 \geq w_2$, then

$$\tilde{f}(w) = w_1 f(\{1\}) + w_2 f(\{2\}|\{1\}) \tag{17.66}$$

  - If $w = (1, 0)/f(\{1\}) = \big(1/f(\{1\}), 0\big)$ then $\tilde{f}(w) = 1$.
  - If $w = (1, 1)/f(\{1, 2\})$ then $\tilde{f}(w) = 1$.

- If $w_1 \leq w_2$, then

$$\tilde{f}(w) = w_2 f(\{2\}) + w_1 f(\{1\}|\{2\}) \tag{17.67}$$

  - If $w = (0, 1)/f(\{2\}) = (0, 1/f(\{2\}))$ then $\tilde{f}(w) = 1$.
  - If $w = (1, 1)/f(\{1, 2\})$ then $\tilde{f}(w) = 1$.

Example: $m = 2$, $E = \{1, 2\}$, contours

- If $w_1 \geq w_2$, then

$$\tilde{f}(w) = w_1 f(\{1\}) + w_2 f(\{2\}|\{1\}) \tag{17.66}$$

  - If $w = (1, 0)/f(\{1\}) = \Big(1/f(\{1\}), 0\Big)$ then $\tilde{f}(w) = 1$.
  - If $w = (1, 1)/f(\{1, 2\})$ then $\tilde{f}(w) = 1$.

- If $w_1 \leq w_2$, then

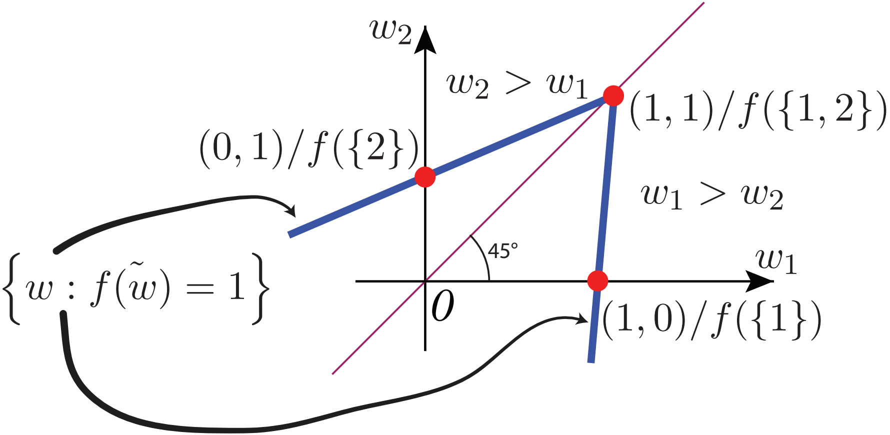
$$\tilde{f}(w) = w_2 f(\{2\}) + w_1 f(\{1\}|\{2\}) \tag{17.67}$$

  - If $w = (0, 1)/f(\{2\}) = (0, 1/f(\{2\}))$ then $\tilde{f}(w) = 1$.
  - If $w = (1, 1)/f(\{1, 2\})$ then $\tilde{f}(w) = 1$.

- Can plot contours of the form $\left\{w \in \mathbb{R}^2 : \tilde{f}(w) = 1\right\}$, particular marked points of form $w = \mathbf{1}_A \times \frac{1}{f(A)}$ for certain A, where $\tilde{f}(w) = 1$.

Example: $m = 2$, $E = \{1, 2\}$

- Contour plot of $m = 2$ Lovász extension (from Bach-2011).



Example: $m = 3$, $E = \{1, 2, 3\}$

- In order to visualize in 3D, we make a few simplifications.

Example: $m = 3$, $E = \{1, 2, 3\}$

- In order to visualize in 3D, we make a few simplifications.
- Consider any submodular f' and $x \in B_{f'}$. Then $f(A) = f'(A) - x(A)$ is submodular

Example: $m = 3$, $E = \{1, 2, 3\}$

- In order to visualize in 3D, we make a few simplifications.
- Consider any submodular f' and $x \in B_{f'}$. Then $f(A) = f'(A) - x(A)$ is submodular, and moreover $f(E) = f'(E) - x(E) = 0$.

Example: $m = 3$, $E = \{1, 2, 3\}$

- In order to visualize in 3D, we make a few simplifications.
- Consider any submodular f' and $x \in B_{f'}$. Then $f(A) = f'(A) - x(A)$ is submodular, and moreover $f(E) = f'(E) - x(E) = 0$.
- Hence, from $\tilde{f}(w + \alpha \mathbf{1}_E) = \tilde{f}(w) + \alpha f(E)$, we have that $\tilde{f}(w + \alpha \mathbf{1}_E) = \tilde{f}(w)$.

Example: $m = 3$, $E = \{1, 2, 3\}$

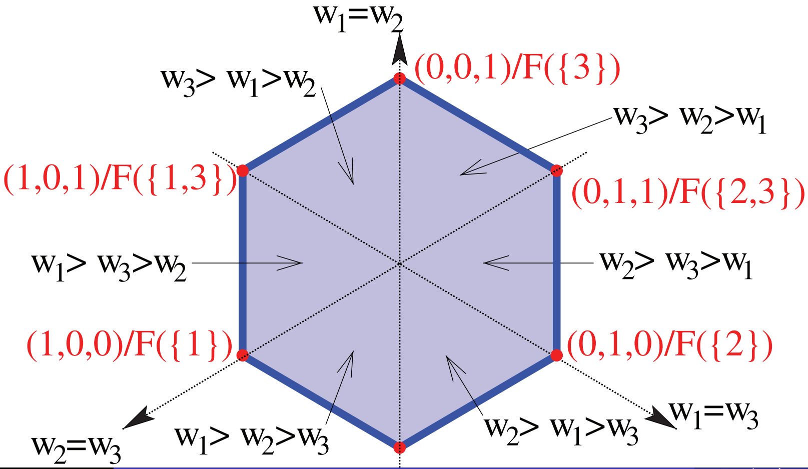
- In order to visualize in 3D, we make a few simplifications.
- Consider any submodular f' and $x \in B_{f'}$. Then $f(A) = f'(A) - x(A)$ is submodular, and moreover $f(E) = f'(E) - x(E) = 0$.
- Hence, from $\tilde{f}(w + \alpha \mathbf{1}_E) = \tilde{f}(w) + \alpha f(E)$, we have that $\tilde{f}(w + \alpha \mathbf{1}_E) = \tilde{f}(w)$.
- Thus, we can look "down" on the contour plot of the Lovász extension, $\left\{w : \tilde{f}(w) = 1\right\}$, from a vantage point right on the line $\{x : x = \alpha \mathbf{1}_E, \alpha > 0\}$ since moving in direction $\mathbf{1}_E$ changes nothing.

Example: $m = 3$, $E = \{1, 2, 3\}$

- Example 1 (from Bach-2011): $f(A) = \mathbf{1}_{|A| \in \{1,2\}}$ $= \min\{|A|, 1\} + \min\{|E \setminus A|, 1\} - 1$ is submodular, and $\tilde{f}(w) = \max_{k \in \{1,2,3\}} w_k - \min_{k \in \{1,2,3\}} w_k$.

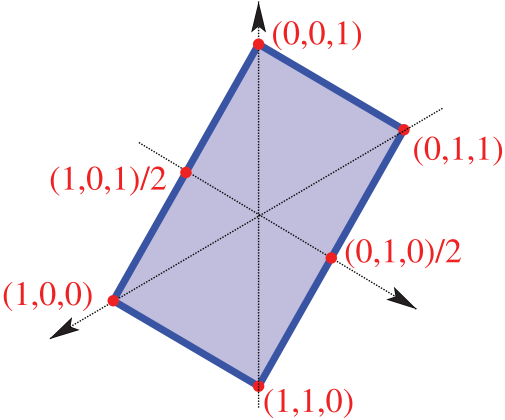
Example: $m = 3,\ E = \{1, 2, 3\}$

- Example 1 (from Bach-2011): $f(A) = \mathbf{1}_{|A| \in \{1,2\}}$ $= \min\{|A|, 1\} + \min\{|E \setminus A|, 1\} - 1$ is submodular, and $\tilde{f}(w) = \max_{k \in \{1,2,3\}} w_k - \min_{k \in \{1,2,3\}} w_k$.



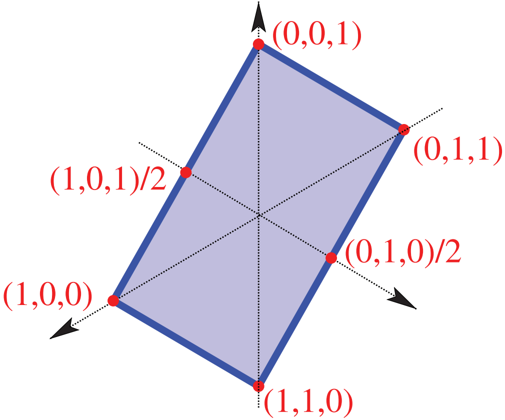
Example: $m = 3,\ E = \{1, 2, 3\}$

- Example 2 (from Bach-2011): $f(A) = |\mathbf{1}_{1 \in A} - \mathbf{1}_{2 \in A}| + |\mathbf{1}_{2 \in A} - \mathbf{1}_{3 \in A}|$



Example: $m = 3$, $E = \{1, 2, 3\}$

- Example 2 (from Bach-2011): $f(A) = |\mathbf{1}_{1 \in A} - \mathbf{1}_{2 \in A}| + |\mathbf{1}_{2 \in A} - \mathbf{1}_{3 \in A}|$
- This gives a "total variation" function for the Lovász extension, with $\tilde{f}(w) = |w_1 - w_2| + |w_2 - w_3|$, a prior to prefer piecewise-constant signals.



Total Variation Example

From “Nonlinear total variation based noise removal algorithms” Rudin, Osher, and Fatemi, 1992. Top left original, bottom right total variation.

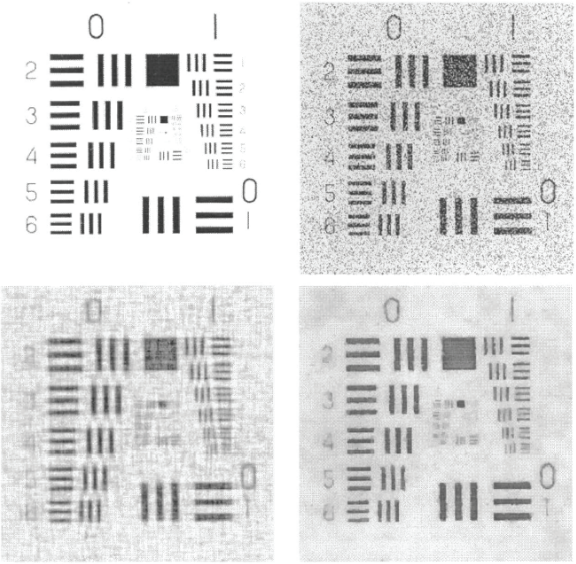


Fig. 3. (a) “Resolution Chart”. (b) Noisy “Resolution Chart”, SNR = 1.0. (c) Wiener filter reconstruction from (b). (d) TV reconstruction from (b).

Example: Lovász extension of concave over modular

- Let $m : E \to \mathbb{R}_+$ be a modular function and define $f(A) = g(m(A))$ where g is concave. Then f is submodular.

Example: Lovász extension of concave over modular

- Let $m : E \to \mathbb{R}_+$ be a modular function and define $f(A) = g(m(A))$ where g is concave. Then f is submodular.
- Let $M_j = \sum_{i=1}^{j} m(e_i)$

Example: Lovász extension of concave over modular

- Let $m : E \to \mathbb{R}_+$ be a modular function and define $f(A) = g(m(A))$ where g is concave. Then f is submodular.
- Let $M_j = \sum_{i=1}^{j} m(e_i)$
- $\tilde{f}(w)$ is given as

$$\tilde{f}(w) = \sum_{i=1}^{m} w(e_i)\big(g(M_i) - g(M_{i-1})\big) \tag{17.68}$$

Example: Lovász extension of concave over modular

- Let $m : E \to \mathbb{R}_+$ be a modular function and define $f(A) = g(m(A))$ where g is concave. Then f is submodular.
- Let $M_j = \sum_{i=1}^{j} m(e_i)$
- $\tilde{f}(w)$ is given as

$$\tilde{f}(w) = \sum_{i=1}^{m} w(e_i)\big(g(M_i) - g(M_{i-1})\big) \tag{17.68}$$

- And if $m(A) = |A|$, we get

$$\tilde{f}(w) = \sum_{i=1}^{m} w(e_i)\big(g(i) - g(i-1)\big) \tag{17.69}$$

Example: Lovász extension and cut functions

- Cut Function: Given a non-negative weighted graph $G = (V, E, m)$ where $m : E \to \mathbb{R}_+$ is a modular function over the edges, we know from Lecture 2 that $f : 2^V \to \mathbb{R}_+$ with $f(X) = m(\Gamma(X))$ where $\Gamma(X) = \{(u, v) | (u, v) \in E, u \in X, v \in V \setminus X\}$ is non-monotone submodular.

Example: Lovász extension and cut functions

- Cut Function: Given a non-negative weighted graph $G = (V, E, m)$ where $m : E \to \mathbb{R}_+$ is a modular function over the edges, we know from Lecture 2 that $f : 2^V \to \mathbb{R}_+$ with $f(X) = m(\Gamma(X))$ where $\Gamma(X) = \{(u, v) | (u, v) \in E, u \in X, v \in V \setminus X\}$ is non-monotone submodular.
- Simple way to write it, with $m_{ij} = m((i, j))$:

$$f(X) = \sum_{i \in X, j \in V \setminus X} m_{ij} \tag{17.70}$$

Example: Lovász extension and cut functions

- Cut Function: Given a non-negative weighted graph $G = (V, E, m)$ where $m : E \to \mathbb{R}_+$ is a modular function over the edges, we know from Lecture 2 that $f : 2^V \to \mathbb{R}_+$ with $f(X) = m(\Gamma(X))$ where $\Gamma(X) = \{(u, v)|(u, v) \in E, u \in X, v \in V \setminus X\}$ is non-monotone submodular.
- Simple way to write it, with $m_{ij} = m((i, j))$:

$$f(X) = \sum_{i \in X, j \in V \setminus X} m_{ij} \tag{17.70}$$

- Exercise: show that Lovász extension of graph cut may be written as:

$$\tilde{f}(w) = \sum_{i,j \in V} m_{ij} \max\{(w_i - w_j), 0\} \tag{17.71}$$

where elements are ordered as usual, $w_1 \geq w_2 \geq \cdots \geq w_n$.

Example: Lovász extension and cut functions

- Cut Function: Given a non-negative weighted graph $G = (V, E, m)$ where $m : E \to \mathbb{R}_+$ is a modular function over the edges, we know from Lecture 2 that $f : 2^V \to \mathbb{R}_+$ with $f(X) = m(\Gamma(X))$ where $\Gamma(X) = \{(u, v) | (u, v) \in E, u \in X, v \in V \setminus X\}$ is non-monotone submodular.
- Simple way to write it, with $m_{ij} = m((i, j))$:

$$f(X) = \sum_{i \in X, j \in V \setminus X} m_{ij} \tag{17.70}$$

- Exercise: show that Lovász extension of graph cut may be written as:

$$\tilde{f}(w) = \sum_{i,j \in V} m_{ij} \max\{(w_i - w_j), 0\} \tag{17.71}$$

  where elements are ordered as usual, $w_1 \geq w_2 \geq \cdots \geq w_n$.
- This is also a form of "total variation"

A few more Lovász extension examples

Some additional submodular functions and their Lovász extensions, where $w(e_1) \geq w(e_2) \geq \cdots \geq w(e_m) \geq 0$. Let $W_k \triangleq \sum_{i=1}^{k} w(e_i)$.

$f(A)$	$\tilde{f}(w)$
$\lvert A\rvert$	$\lVert w\rVert_1$
$\min(\lvert A\rvert, 1)$	$\lVert w\rVert_\infty$
$\min(\lvert A\rvert, 1) - \max(\lvert A\rvert - m + 1, 0)$	$\lVert w\rVert_\infty - \min_i w_i$
$\min(\lvert A\rvert, k)$	W_k
$\min(\lvert A\rvert, k) - \max(\lvert A\rvert - (n-k) + 1, 1)$	$2W_k - W_m$
$\min(\lvert A\rvert, \lvert E \setminus A\rvert)$	$2W_{\lfloor m/2 \rfloor} - W_m$

(thanks to K. Narayanan).

Supervised And Unsupervised Machine Learning

- Given training data $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^m$ with $(x_i, y_i) \in \mathbb{R}^n \times \mathbb{R}$, perform the following risk minimization problem:

$$\min_{w \in \mathbb{R}^n} \frac{1}{m} \sum_{i=1}^{m} \ell(y_i, w^\intercal x_i) + \lambda \Omega(w), \tag{17.72}$$

where $\ell(\cdot)$ is a loss function (e.g., squared error) and $\Omega(w)$ is a norm.

Supervised And Unsupervised Machine Learning

- Given training data $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^m$ with $(x_i, y_i) \in \mathbb{R}^n \times \mathbb{R}$, perform the following risk minimization problem:

$$\min_{w \in \mathbb{R}^n} \frac{1}{m} \sum_{i=1}^m \ell(y_i, w^\intercal x_i) + \lambda \Omega(w), \tag{17.72}$$

where $\ell(\cdot)$ is a loss function (e.g., squared error) and $\Omega(w)$ is a norm.

- When data has multiple responses $(x_i, y_i) \in \mathbb{R}^n \times \mathbb{R}^k$, learning becomes:

$$\min_{w^1, \dots, w^k \in \mathbb{R}^n} \sum_{j=1}^k \frac{1}{m} \sum_{i=1}^m \ell(y_i^k, (w^k)^\intercal x_i) + \lambda \Omega(w^k), \tag{17.73}$$

Supervised And Unsupervised Machine Learning

- Given training data $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^m$ with $(x_i, y_i) \in \mathbb{R}^n \times \mathbb{R}$, perform the following risk minimization problem:

$$\min_{w \in \mathbb{R}^n} \frac{1}{m} \sum_{i=1}^m \ell(y_i, w^\intercal x_i) + \lambda \Omega(w), \tag{17.72}$$

where $\ell(\cdot)$ is a loss function (e.g., squared error) and $\Omega(w)$ is a norm.

- When data has multiple responses $(x_i, y_i) \in \mathbb{R}^n \times \mathbb{R}^k$, learning becomes:

$$\min_{w^1, \dots, w^k \in \mathbb{R}^n} \sum_{j=1}^k \frac{1}{m} \sum_{i=1}^m \ell(y_i^k, (w^k)^\intercal x_i) + \lambda \Omega(w^k), \tag{17.73}$$

- When data has multiple responses only that are observed, $(y_i) \in R^k$ we get dictionary learning (Krause & Guestrin, Das & Kempe):

$$\min_{x_1, \dots, x_m} \min_{w^1, \dots, w^k \in \mathbb{R}^n} \sum_{j=1}^k \frac{1}{m} \sum_{i=1}^m \ell(y_i^k, (w^k)^\intercal x_i) + \lambda \Omega(w^k), \tag{17.74}$$

Norms, sparse norms, and computer vision

- Common norms include p-norm $\Omega(w) = \|w\|_p = \left(\sum_{i=1}^{p} w_i^p\right)^{1/p}$

Norms, sparse norms, and computer vision

- Common norms include p-norm $\Omega(w) = \|w\|_p = \left(\sum_{i=1}^{p} w_i^p\right)^{1/p}$
- 1-norm promotes sparsity (prefer solutions with zero entries).

Norms, sparse norms, and computer vision

- Common norms include p-norm $\Omega(w) = \|w\|_p = \left(\sum_{i=1}^p w_i^p\right)^{1/p}$
- 1-norm promotes sparsity (prefer solutions with zero entries).
- Image denoising, total variation is useful, norm takes form:

$$\Omega(w) = \sum_{i=2}^{N} |w_i - w_{i-1}| \tag{17.75}$$

Norms, sparse norms, and computer vision

- Common norms include p-norm $\Omega(w) = \|w\|_p = \left(\sum_{i=1}^p w_i^p\right)^{1/p}$
- 1-norm promotes sparsity (prefer solutions with zero entries).
- Image denoising, total variation is useful, norm takes form:

$$\Omega(w) = \sum_{i=2}^{N} |w_i - w_{i-1}| \tag{17.75}$$

- Points of difference should be "sparse" (frequently zero).

(Rodriguez, 2009)

Submodular parameterization of a sparse convex norm

- Prefer convex norms since they can be solved.

Submodular parameterization of a sparse convex norm

- Prefer convex norms since they can be solved.
- For $w \in \mathbb{R}^V$, $\text{supp}(w) \in \{0,1\}^V$ has $\text{supp}(w)(v) = 1$ iff $w(v) > 0$

Submodular parameterization of a sparse convex norm

- Prefer convex norms since they can be solved.
- For $w \in \mathbb{R}^V$, $\text{supp}(w) \in \{0,1\}^V$ has $\text{supp}(w)(v) = 1$ iff $w(v) > 0$
- Desirable sparse norm: count the non-zeros, $\|w\|_0 = \text{supp}(w)$.

Submodular parameterization of a sparse convex norm

- Prefer convex norms since they can be solved.
- For $w \in \mathbb{R}^V$, $\text{supp}(w) \in \{0,1\}^V$ has $\text{supp}(w)(v) = 1$ iff $w(v) > 0$
- Desirable sparse norm: count the non-zeros, $\|w\|_0 = \text{supp}(w)$.
- Using $\Omega(w) = \|w\|_0$ is NP-hard, instead we often optimize tightest convex relaxation, $\|w\|$ which is the convex envelope.

Submodular parameterization of a sparse convex norm

- Prefer convex norms since they can be solved.
- For $w \in \mathbb{R}^V$, $\text{supp}(w) \in \{0,1\}^V$ has $\text{supp}(w)(v) = 1$ iff $w(v) > 0$
- Desirable sparse norm: count the non-zeros, $\|w\|_0 = \text{supp}(w)$.
- Using $\Omega(w) = \|w\|_0$ is NP-hard, instead we often optimize tightest convex relaxation, $\|w\|$ which is the convex envelope.
- With $\|w\|_0$ or its relaxation, each non-zero element has equal degree of penalty. Penalties do not interact.

Submodular parameterization of a sparse convex norm

- Prefer convex norms since they can be solved.
- For $w \in \mathbb{R}^V$, $\text{supp}(w) \in \{0,1\}^V$ has $\text{supp}(w)(v) = 1$ iff $w(v) > 0$
- Desirable sparse norm: count the non-zeros, $\|w\|_0 = \text{supp}(w)$.
- Using $\Omega(w) = \|w\|_0$ is NP-hard, instead we often optimize tightest convex relaxation, $\|w\|$ which is the convex envelope.
- With $\|w\|_0$ or its relaxation, each non-zero element has equal degree of penalty. Penalties do not interact.
- Given submodular function $f : 2^V \to \mathbb{R}_+$, $f(\text{supp}(w))$ measures the "complexity" of the non-zero pattern of w; can have more non-zero values if they cooperate (via f) with other non-zero values.

Submodular parameterization of a sparse convex norm

- Prefer convex norms since they can be solved.
- For $w \in \mathbb{R}^V$, $\text{supp}(w) \in \{0,1\}^V$ has $\text{supp}(w)(v) = 1$ iff $w(v) > 0$
- Desirable sparse norm: count the non-zeros, $\|w\|_0 = \text{supp}(w)$.
- Using $\Omega(w) = \|w\|_0$ is NP-hard, instead we often optimize tightest convex relaxation, $\|w\|$ which is the convex envelope.
- With $\|w\|_0$ or its relaxation, each non-zero element has equal degree of penalty. Penalties do not interact.
- Given submodular function $f : 2^V \to \mathbb{R}_+$, $f(\text{supp}(w))$ measures the "complexity" of the non-zero pattern of w; can have more non-zero values if they cooperate (via f) with other non-zero values.
- $f(\text{supp}(w))$ is hard to optimize, but it's convex envelope $\tilde{f}(|w|)$ (i.e., largest convex under-estimator of $f(\text{supp}(w))$) is obtained via the Lovász-extension $\tilde{f}$ of f (Bach 2010).

Submodular parameterization of a sparse convex norm

- Prefer convex norms since they can be solved.
- For $w \in \mathbb{R}^V$, $\text{supp}(w) \in \{0,1\}^V$ has $\text{supp}(w)(v) = 1$ iff $w(v) > 0$
- Desirable sparse norm: count the non-zeros, $\|w\|_0 = \text{supp}(w)$.
- Using $\Omega(w) = \|w\|_0$ is NP-hard, instead we often optimize tightest convex relaxation, $\|w\|$ which is the convex envelope.
- With $\|w\|_0$ or its relaxation, each non-zero element has equal degree of penalty. Penalties do not interact.
- Given submodular function $f : 2^V \to \mathbb{R}_+$, $f(\text{supp}(w))$ measures the "complexity" of the non-zero pattern of w; can have more non-zero values if they cooperate (via f) with other non-zero values.
- $f(\text{supp}(w))$ is hard to optimize, but it's convex envelope $\tilde{f}(|w|)$ (i.e., largest convex under-estimator of $f(\text{supp}(w))$) is obtained via the Lovász-extension $\tilde{f}$ of f (Bach 2010).
- Submodular functions thus parameterize structured convex sparse norms via the Lovász-extension!

Submodular parameterization of a sparse convex norm

- Prefer convex norms since they can be solved.
- For $w \in \mathbb{R}^V$, $\text{supp}(w) \in \{0,1\}^V$ has $\text{supp}(w)(v) = 1$ iff $w(v) > 0$
- Desirable sparse norm: count the non-zeros, $\|w\|_0 = \text{supp}(w)$.
- Using $\Omega(w) = \|w\|_0$ is NP-hard, instead we often optimize tightest convex relaxation, $\|w\|$ which is the convex envelope.
- With $\|w\|_0$ or its relaxation, each non-zero element has equal degree of penalty. Penalties do not interact.
- Given submodular function $f : 2^V \to \mathbb{R}_+$, $f(\text{supp}(w))$ measures the "complexity" of the non-zero pattern of w; can have more non-zero values if they cooperate (via f) with other non-zero values.
- $f(\text{supp}(w))$ is hard to optimize, but it's convex envelope $\tilde{f}(|w|)$ (i.e., largest convex under-estimator of $f(\text{supp}(w))$) is obtained via the Lovász-extension $\tilde{f}$ of f (Bach 2010).
- Submodular functions thus parameterize structured convex sparse norms via the Lovász-extension!
- Ex: total variation is Lovász-ext. of graph cut, but $\exists$ many more!

Lovász extension and norms

- Using Lovász extension to define various norms of the form $\|w\|_{\tilde{f}} = \tilde{f}(|w|)$, renders the functionf symmetric about all orthants (i.e., $\|w\|_{\tilde{f}} = \|b \odot w\|_{\tilde{f}}$ where $b \in \{-1, 1\}^m$ and $\odot$ is element-wise multiplication).

Lovász extension and norms

- Using Lovász extension to define various norms of the form $\|w\|_{\tilde{f}} = \tilde{f}(|w|)$, renders the functionf symmetric about all orthants (i.e., $\|w\|_{\tilde{f}} = \|b \odot w\|_{\tilde{f}}$ where $b \in \{-1, 1\}^m$ and $\odot$ is element-wise multiplication).
- Simple example. The Lovász extension of the modular function $f(A) = |A|$ is the ℓ_1 norm, and the Lovász extension of the modular function $f(A) = m(A)$ is the weighted ℓ_1 norm.

Lovász extension and norms

- Using Lovász extension to define various norms of the form $\|w\|_{\tilde{f}} = \tilde{f}(|w|)$, renders the functionf symmetric about all orthants (i.e., $\|w\|_{\tilde{f}} = \|b \odot w\|_{\tilde{f}}$ where $b \in \{-1, 1\}^m$ and $\odot$ is element-wise multiplication).
- Simple example. The Lovász extension of the modular function $f(A) = |A|$ is the ℓ_1 norm, and the Lovász extension of the modular function $f(A) = m(A)$ is the weighted ℓ_1 norm.
- With more general submodular functions, one can generate a large and interesting variety of norms, all of which have polyhedral contours (unlike, say, something like the ℓ_2 norm).

Lovász extension and norms

- Using Lovász extension to define various norms of the form $\|w\|_{\tilde{f}} = \tilde{f}(|w|)$, renders the functionf symmetric about all orthants (i.e., $\|w\|_{\tilde{f}} = \|b \odot w\|_{\tilde{f}}$ where $b \in \{-1, 1\}^m$ and $\odot$ is element-wise multiplication).
- Simple example. The Lovász extension of the modular function $f(A) = |A|$ is the ℓ_1 norm, and the Lovász extension of the modular function $f(A) = m(A)$ is the weighted ℓ_1 norm.
- With more general submodular functions, one can generate a large and interesting variety of norms, all of which have polyhedral contours (unlike, say, something like the ℓ_2 norm).
- Hence, not all norms come from the Lovász extension of some submodular function.

Lovász extension and norms

- Using Lovász extension to define various norms of the form $\|w\|_{\tilde{f}} = \tilde{f}(|w|)$, renders the functionf symmetric about all orthants (i.e., $\|w\|_{\tilde{f}} = \|b \odot w\|_{\tilde{f}}$ where $b \in \{-1, 1\}^m$ and $\odot$ is element-wise multiplication).
- Simple example. The Lovász extension of the modular function $f(A) = |A|$ is the ℓ_1 norm, and the Lovász extension of the modular function $f(A) = m(A)$ is the weighted ℓ_1 norm.
- With more general submodular functions, one can generate a large and interesting variety of norms, all of which have polyhedral contours (unlike, say, something like the ℓ_2 norm).
- Hence, not all norms come from the Lovász extension of some submodular function.
- Similarly, not all convex functions are the Lovász extension of some submodular function.

Lovász extension and norms

- Using Lovász extension to define various norms of the form $\|w\|_{\tilde{f}} = \tilde{f}(|w|)$, renders the functionf symmetric about all orthants (i.e., $\|w\|_{\tilde{f}} = \|b \odot w\|_{\tilde{f}}$ where $b \in \{-1,1\}^m$ and $\odot$ is element-wise multiplication).
- Simple example. The Lovász extension of the modular function $f(A) = |A|$ is the ℓ_1 norm, and the Lovász extension of the modular function $f(A) = m(A)$ is the weighted ℓ_1 norm.
- With more general submodular functions, one can generate a large and interesting variety of norms, all of which have polyhedral contours (unlike, say, something like the ℓ_2 norm).
- Hence, not all norms come from the Lovász extension of some submodular function.
- Similarly, not all convex functions are the Lovász extension of some submodular function.
- Bach-2011 has a complete discussion of this.