Submodular Functions, Optimization, and Applications to Machine Learning

— Spring Quarter, Lecture 16 —

http://www.ee.washington.edu/people/faculty/bilmes/classes/ee563_spring_2018/

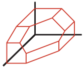
Prof. Jeff Bilmes

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

May 21st, 2018

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

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



Announcements, Assignments, and Reminders

4 more lectures

- Next homework will be posted tonight.
- Rest of the quarter. One more longish homework.
- Take home final exam (like a long homework).
- As always, if you have any questions about anything, please ask then via our discussion board (https://canvas.uw.edu/courses/1216339/discussion_topics). Can meet at odd hours via zoom (send message on canvas to schedule time to chat).

Class Road Map - EE563

- L1(3/26): Motivation, Applications, & Basic Definitions,
- L2(3/28): Machine Learning Apps (diversity, complexity, parameter, learning target, surrogate).
- L3(4/2): Info theory exs, more apps, definitions, graph/combinatorial examples
- L4(4/4): Graph and Combinatorial Examples, Matrix Rank, Examples and Properties, visualizations
- L5(4/9): More Examples/Properties/ Other Submodular Defs., Independence,
- L6(4/11): Matroids, Matroid Examples, Matroid Rank, Partition/Laminar Matroids
- L7(4/16): Laminar Matroids, System of Distinct Reps, Transversals, Transversal Matroid, Matroid Representation, Dual Matroids
- L8(4/18): Dual Matroids, Other Matroid Properties, Combinatorial Geometries, Matroids and Greedy.
- L9(4/23): Polyhedra, Matroid Polytopes, Matroids → Polymatroids
- L10(4/29): Matroids → Polymatroids, Polymatroids, Polymatroids and Greedy,
- L11(4/30): Polymatroids, Polymatroids and Greedy
- L12(5/2): Polymatroids and Greedy, Extreme Points, Cardinality Constrained Maximization
- L13(5/7): Constrained Submodular Maximization
- L14(5/9): Submodular Max w. Other Constraints, Cont. Extensions, Lovasz Extension
- L15(5/14): Cont. Extensions, Lovasz Extension, Choquet Integration, Properties
- L16(5/16): More Lovasz extension, Choquet, defs/props, examples, multiliear extension
- L17(5/21):
- L18(5/23):
- L–(5/28): Memorial Day (holiday).
- L19(5/30):
- L21(6/4): Final Presentations maximization.

Last day of instruction, June 1st. Finals Week: June 2-8, 2018.

Convex Closure of Discrete Set Functions

- Given set function $f : 2^V \to \mathbb{R}$, an arbitrary (i.e., not necessarily submodular nor supermodular) set function, define a function $\check{f} : [0,1]^V \to \mathbb{R}$, as

$$\check{f}(x) = \min_{p \in \triangle^n(x)} \sum_{S \subseteq V} p_S f(S) \tag{16.1}$$

where $\triangle^n(x) =$
$\left\{ p \in \mathbb{R}^{2^n} : \sum_{S \subseteq V} p_S = 1,\ p_S \geq 0 \forall S \subseteq V,\ \&\ \sum_{S \subseteq V} p_S \mathbf{1}_S = x \right\}$

- Hence, $\triangle^n(x)$ is the set of all probability distributions over the 2^n vertices of the hypercube, and where the expected value of the characteristic vectors of those points is equal to x, i.e., for any $p \in \triangle^n(x)$, $E_{S \sim p}(\mathbf{1}_S) = \sum_{S \subseteq V} p_S \mathbf{1}_S = x$.
- Hence, $\check{f}(x) = \min_{p \in \triangle^n(x)} E_{S \sim p}[f(S)]$
- Note, this is not (necessarily) the Lovász extension, rather this is **a** convex extension.

Convex Closure of Discrete Set Functions

- Given, $\check{f}(x) = \min_{p \in \triangle^n(x)} E_{S \sim p}[f(S)]$, we can show:
  1. That $\check{f}$ is tight (i.e., $\forall S \subseteq V$, we have $\check{f}(\mathbf{1}_S) = f(S)$).
  2. That $\check{f}$ is convex (and consequently, that any arbitrary set function has a tight convex extension).
  3. That the convex closure $\check{f}$ is the convex envelope of the function defined only on the hypercube vertices, and that takes value $f(S)$ at $\mathbf{1}_S$.
  4. The definition of the Lovász extension of a set function, and that $\check{f}$ is the Lovász extension iff f is submodular.

A continuous extension of submodular f

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

$$\breve{f}(w) = \max(wx : x \in B_f) \tag{16.12}$$

$$= \sum_{i=1}^{m} w(e_i) f(e_i | E_{i-1}) = \sum_{i=1}^{m} w(e_i) x(e_i) \tag{16.13}$$

$$= \sum_{i=1}^{m} w(e_i)(f(E_i) - f(E_{i-1})) \tag{16.14}$$

$$= w(e_m) f(E_m) + \sum_{i=1}^{m-1} (w(e_i) - w(e_{i+1})) f(E_i) \tag{16.15}$$

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

A continuous extension of submodular f

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

$$\breve{f}(w) = \max(wx : x \in B_f) \tag{16.12}$$

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

$$\breve{f}(w) = w(e_m)f(E_m) + \sum_{i=1}^{m-1}(w(e_i) - w(e_{i+1}))f(E_i) \tag{16.13}$$

$$= \sum_{i=1}^{m} \lambda_i f(E_i) \tag{16.14}$$

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

- Convex analysis $\Rightarrow$ $\breve{f}(w) = \max(wx : x \in P)$ is always convex in w for any set $P \subseteq R^E$, since a maximum of a set of linear functions (true even when f is not submodular or P is not itself 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 $\breve{f}(\mathbf{1}_A) = f(A)$, $\forall A$, in this way where

$$\breve{f}(w) = \sum_{i=1}^{m} \lambda_i f(E_i) \tag{16.21}$$

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

$$\text{for } i \in \{1, \ldots, m\}, \quad \lambda_i = \begin{cases} w(e_i) - w(e_{i+1}) & \text{if } i < m \\ w(e_m) & \text{if } i = m \end{cases} \tag{16.22}$$

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.
- $\breve{f}(w) = \sum_{i=1}^{m} \lambda_i f(E_i)$ is the associated interpolation of the values of f at sets corresponding to each hypercube vertex.
- This extension is called the Lovász extension!

Summary: comparison of the two extension forms

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

$$\breve{f}(w) = \max(wx : x \in B_f) = \sum_{i=1}^{m} \lambda_i f(E_i) \tag{16.25}$$

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 $\breve{f}$ having the form

$$\breve{f}(w) = \sum_{i=1}^{m} \lambda_i f(E_i) \tag{16.26}$$

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

Lovász Extension, Submodularity and Convexity

Theorem 16.2.5

A function $f : 2^E \to \mathbb{R}$ is submodular iff its Lovász extension $\breve{f}$ 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 $\breve{f}(w) = \max \{wx : x \in P_f\}$, and thus is convex.
- Conversely, suppose the Lovász extension $\breve{f}(w) = \sum_i \lambda_i f(E_i)$ 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 $\breve{f}(\alpha w) = \alpha \breve{f}(w)$ for any $\alpha \in \mathbb{R}_+$. I.e., f is a positively homogeneous convex function.

...

Lovász ext. vs. the concave closure of submodular function

Theorem 16.2.5

Let $\breve{f}(w) = \max(wx : x \in B_f) = \sum_{i=1}^m \lambda_i f(E_i)$ *be the Lovász extension and* $\breve{f}(w) = \min_{p \in \triangle^n(w)} E_{S \sim p}[f(S)]$ *be the convex closure. Then* $\breve{f}$ *and* $\breve{f}$ *coincide iff* f *is submodular.*

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

Proof.

- Assume f is submodular.
- Given x, let p^x be an achieving argmin in $\breve{f}(x)$ that also maximizes $\sum_S p^x_S |S|^2$.
- Suppose $\exists A, B \subseteq V$ that are crossing (i.e., $A \not\subseteq B$, $B \not\subseteq A$) and positive and w.l.o.g., $p^x_A \geq p^x_B > 0$.
- Then we may update p^x as follows:

$$\bar{p}^x_A \leftarrow p^x_A - p^x_B \qquad \bar{p}^x_B \leftarrow p^x_B - p^x_B \tag{16.34}$$

$$\bar{p}^x_{A \cup B} \leftarrow p^x_{A \cup B} + p^x_B \qquad \bar{p}^x_{A \cap B} \leftarrow p^x_{A \cap B} + p^x_B \tag{16.35}$$

and by submodularity, this does not increase $\sum_S p^x_S f(S)$.

...

Lovász ext. vs. the concave closure of submodular function

. . . proof cont.

- Next, assume f is not submodular. We must show that the Lovász extension $\breve{f}(x)$ and the concave closure $\breve{f}(x)$ need not coincide.

Lovász ext. vs. the concave closure of submodular function

. . . proof cont.

- Next, assume f is not submodular. We must show that the Lovász extension $\breve{f}(x)$ and the concave closure $\check{f}(x)$ need not coincide.
- Since f is not submodular, $\exists S$ and $i, j \notin S$ such that $f(S) + f(S + i + j) > f(S + i) + f(S + j)$, a strict violation of submodularity.

Lovász ext. vs. the concave closure of submodular function

. . . proof cont.

- Next, assume f is not submodular. We must show that the Lovász extension $\breve{f}(x)$ and the concave closure $\check{f}(x)$ need not coincide.
- Since f is not submodular, $\exists S$ and $i, j \notin S$ such that $f(S) + f(S + i + j) > f(S + i) + f(S + j)$, a strict violation of submodularity.
- Consider $x = \mathbf{1}_S + \frac{1}{2}\mathbf{1}_{\{i,j\}}$.

Lovász ext. vs. the concave closure of submodular function

. . . proof cont.

- Next, assume f is not submodular. We must show that the Lovász extension $\breve{f}(x)$ and the concave closure $\check{f}(x)$ need not coincide.
- Since f is not submodular, $\exists S$ and $i, j \notin S$ such that $f(S) + f(S + i + j) > f(S + i) + f(S + j)$, a strict violation of submodularity.
- Consider $x = \mathbf{1}_S + \frac{1}{2}\mathbf{1}_{\{i,j\}}$.
- Then L.E. has $\breve{f}(x) = \frac{1}{2}f(S) + \frac{1}{2}f(S + i + j)$ and this p^x is feasible for $\check{f}(x)$ with $p^x_S = 1/2$ and $p^x_{S+i+j} = 1/2$.

[1 1 1 ... 1 ½ ½ 0 0 ... 0] (S, {i,j})

$p^x = \{0 0 0 \frac{1}{2} \; \frac{1}{2} \cdots 0\}$

$p^x \in \Delta(x)$

Lovász ext. vs. the concave closure of submodular function

. . . proof cont.

- Next, assume f is not submodular. We must show that the Lovász extension $\breve{f}(x)$ and the concave closure $\breve{f}(x)$ need not coincide.
- Since f is not submodular, $\exists S$ and $i, j \notin S$ such that $f(S) + f(S+i+j) > f(S+i) + f(S+j)$, a strict violation of submodularity.
- Consider $x = \mathbf{1}_S + \frac{1}{2}\mathbf{1}_{\{i,j\}}$.
- Then L.E. has $\breve{f}(x) = \frac{1}{2}f(S) + \frac{1}{2}f(S+i+j)$ and this p^x is feasible for $\breve{f}(x)$ with $p^x_S = 1/2$ and $p^x_{S+i+j} = 1/2$.

  $p^x \in \Delta(x)$
- An alternate feasible distribution for $\breve{f}(x)$ in the convex closure is $\bar{p}^x_{S+i} = \bar{p}^x_{S+j} = 1/2$.

  $\sum_{A \subseteq V} p^x_A \cdot \mathbf{1}_A = x$

  $\bar{p}^x \in \Delta(x)$

Lovász ext. vs. the concave closure of submodular function

. . . proof cont.

$f(x) = x + e^{-x}$

- Next, assume f is not submodular. We must show that the Lovász extension $\breve{f}(x)$ and the concave closure $\check{f}(x)$ need not coincide.
- Since f is not submodular, $\exists S$ and $i, j \notin S$ such that $f(S) + f(S + i + j) > f(S + i) + f(S + j)$, a strict violation of submodularity.
- Consider $x = \mathbf{1}_S + \frac{1}{2}\mathbf{1}_{\{i,j\}}$.

Note: here p^x is not nec. the achieving argument in $\check{f}$.

- Then L.E. has $\breve{f}(x) = \frac{1}{2}f(S) + \frac{1}{2}f(S + i + j)$ and this p^x is feasible for $\check{f}(x)$ with $p^x_S = 1/2$ and $p^x_{S+i+j} = 1/2$.
- An alternate feasible distribution for $\check{f}(x)$ in the convex closure is $\bar{p}^x_{S+i} = \bar{p}^x_{S+j} = 1/2$.
- This gives

$\sum_{A \subseteq V} \bar{p}^x_A \cdot f(A)$

$$\check{f}(x) \leq \frac{1}{2}[f(S + i) + f(S + j)] < \breve{f}(x) \qquad (16.1)$$

meaning $\check{f}(x) \neq \breve{f}(x)$.

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

$$\int_X f du = \sup I_X(s) \tag{16.2}$$

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 the weighted average of x as:

$$\text{WAVG}(x) = \sum_{e \in E} x(e) w(e) \tag{16.3}$$

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 the weighted average of x as:

$$\text{WAVG}(x) = \sum_{e \in E} x(e)w(e) \tag{16.3}$$

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

$$\text{WAVG}(\mathbf{1}_e) = w(e) \tag{16.4}$$

Integration, Aggregation, and Weighted Averages

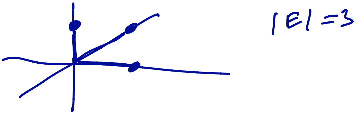
- In finite discrete spaces, Lebesgue integration is just a weighted average, and can be seen as an aggregation function.
- I.e., given a weight vector $w \in [0,1]^E$ for some finite ground set E, then for any $x \in \mathbb{R}^E$ we have the weighted average of x as:

$$\text{WAVG}(x) = \sum_{e \in E} x(e) w(e) \tag{16.3}$$

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

$$\text{WAVG}(\mathbf{1}_e) = w(e) \tag{16.4}$$

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 the weighted average of x as:

$$\text{WAVG}(x) = \sum_{e \in E} x(e) w(e) \tag{16.3}$$

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

$$\text{WAVG}(\mathbf{1}_e) = w(e) \tag{16.4}$$

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

$$\text{WAVG}(x) = \sum_{e \in E} x(e) w(e) = \sum_{e \in E} x(e) \text{WAVG}(\mathbf{1}_e) \tag{16.5}$$

Integration, Aggregation, and Weighted Averages

$$\text{WAVG}(x) = \sum_{e \in E} x(e)w(e) \tag{16.6}$$

- Clearly, WAVG function is linear in weights w, in the argument x, and is homogeneous. That is, for all $w, w_1, w_2, x, x_1, x_2 \in \mathbb{R}^E$ and $\alpha \in \mathbb{R}$,

$$\text{WAVG}_{w_1+w_2}(x) = \text{WAVG}_{w_1}(x) + \text{WAVG}_{w_2}(x), \tag{16.7}$$

$$\text{WAVG}_w(x_1 + x_2) = \text{WAVG}_w(x_1) + \text{WAVG}_w(x_2), \tag{16.8}$$

and is homogeneous, $\forall \alpha \in \mathbb{R}$,

$$\text{WAVG}(\alpha x) = \alpha \text{WAVG}(x). \tag{16.9}$$

Integration, Aggregation, and Weighted Averages

$\breve{f}(x) = \sum_i \lambda_i f(E_i)$

$$\text{WAVG}(x) = \sum_{e \in E} x(e)w(e) \tag{16.6}$$

- Clearly, WAVG function is linear in weights w, in the argument x, and is homogeneous. That is, for all $w, w_1, w_2, x, x_1, x_2 \in \mathbb{R}^E$ and $\alpha \in \mathbb{R}$,

$$\text{WAVG}_{w_1+w_2}(x) = \text{WAVG}_{w_1}(x) + \text{WAVG}_{w_2}(x), \tag{16.7}$$

$$\text{WAVG}_w(x_1 + x_2) = \text{WAVG}_w(x_1) + \text{WAVG}_w(x_2), \tag{16.8}$$

and is homogeneous, $\forall \alpha \in \mathbb{R}$,

$f_{12} = f_1 + f_2$

$$\text{WAVG}(\alpha x) = \alpha \text{WAVG}(x). \tag{16.9}$$

$\breve{f}_{12}(x) = \breve{f}_1(x) + \breve{f}_2(x)$

- How related? The Lovász extension $\breve{f}(x)$ is still linear in "weights" (i.e., the submodular function f), but will not be linear in x and will only be positively homogeneous (for $\alpha \geq 0$).

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 $A \subseteq E$):

$$\mathsf{AG}(\mathbf{1}_A) = w_A \tag{16.10}$$

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 $A \subseteq E$):

$$\text{AG}(\mathbf{1}_A) = w_A \in \mathbb{R} \tag{16.10}$$

$|\{w_A\}_{A \subseteq E}| = 2^{|E|}$

- 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:

$$\text{AG}(x) = \sum_{A \subseteq E} x(A) w_A = \sum_{A \subseteq E} x(A) \text{AG}(\mathbf{1}_A) \tag{16.11}$$

$i \in E$ $\quad$ $x(A) = \sum_{a \in A} x(a)$

Integration, Aggregation, and Weighted Averages

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

$$\text{AG}(\mathbf{1}_A) = w_A \tag{16.10}$$

- 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:

$$\text{AG}(x) = \sum_{A \subseteq E} x(A) w_A = \sum_{A \subseteq E} x(A) \text{AG}(\mathbf{1}_A) \tag{16.11}$$

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

$$\text{WAVG}_{w'}(x) = \text{AG}_w(x) \tag{16.12}$$

Integration, Aggregation, and Weighted Averages

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

$$\text{AG}(\mathbf{1}_A) = w_A \tag{16.10}$$

- 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:

$$\text{AG}(x) = \sum_{A \subseteq E} x(A) w_A = \sum_{A \subseteq E} x(A) \text{AG}(\mathbf{1}_A) \tag{16.11}$$

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

$$\text{WAVG}_{w'}(x) = \text{AG}_w(x) \tag{16.12}$$

- 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 16.4.1

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

$$C_f(w) = \sum_{i=1}^{m} (w_{e_i} - w_{e_{i+1}}) f(E_i) \tag{16.13}$$

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

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

$$C_f(w) = \sum_{i=1}^{m} w(e_i)(f(E_i) - f(E_{i-1})) \tag{16.14}$$

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

Choquet integral

Definition 16.4.1

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

$$C_f(w) = \sum_{i=1}^{m} (w_{e_i} - w_{e_{i+1}}) f(E_i) \tag{16.13}$$

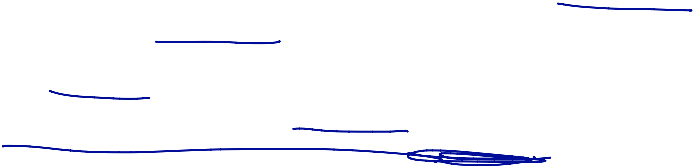
where in the sum, we have sorted and renamed the elements of E so that $w_{e_1} \geq w_{e_2} \geq \dots \geq w_{e_m} \geq w_{e_{m+1}} \triangleq 0$, and where $E_i = \{e_1, e_2, \dots, e_i\}$.

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

$$\sum_{k=m}^{n} a_k b_k = \sum_{k=m}^{n} A_k (b_k - b_{k+1}) + A_n b_{n+1} - A_{m-1} b_m \tag{16.15}$$

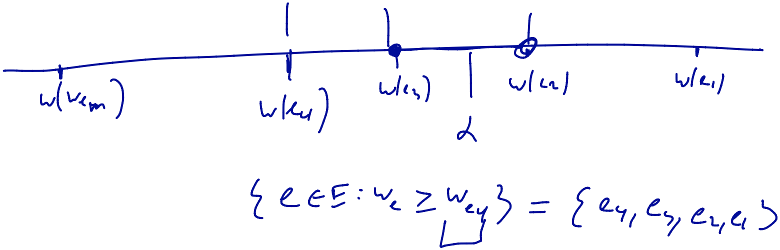
The "integral" in the Choquet integral

- 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 $E_i = \{e_1, e_2, \ldots, e_i\} = \{e \in E : w_e \geq w_{e_i}\}$.

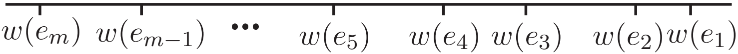


The "integral" in the Choquet integral

- 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 $E_i = \{e_1, e_2, \ldots, e_i\} = \{e \in E : w_e > \alpha\}$.

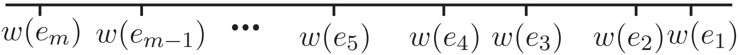
The "integral" in the Choquet integral

- 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 $E_i = \{e_1, e_2, \ldots, e_i\} = \{e \in E : w_e \geq w_{e_i}\}$.
- For any $w_{e_i} > \alpha \geq w_{e_{i+1}}$ we also have $E_i = \{e_1, e_2, \ldots, e_i\} = \{e \in E : w_e > \alpha\}$.
- Can segment 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 $E_i = \{e_1, e_2, \ldots, e_i\} = \{e \in E : w_e > \alpha\}$.
- Can segment real-axis at boundary points w_{e_i}, right most is w_{e_1}.



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

$$F_i(\alpha) = f(\{e \in E : w_e > \alpha\}) = f(E_i) \tag{16.16}$$

The "integral" in the Choquet integral

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

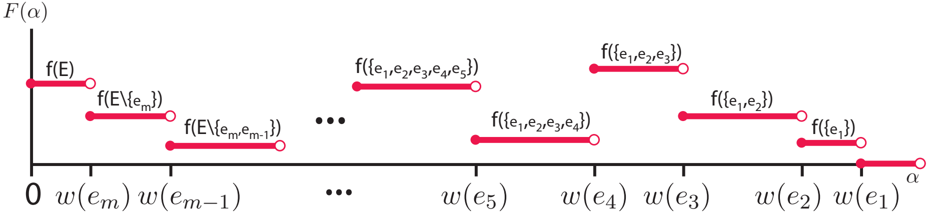
$$F(\alpha) = \begin{cases} f(E) & \text{if } 0 \leq \alpha < w_m \\ f(\{e \in E : w_e > \alpha\}) & \text{if } w_{e_{i+1}} \leq \alpha < w_{e_i},\ i \in \{1, \dots, m-1\} \\ 0\ (= f(\emptyset)) & \text{if } w_1 < \alpha \end{cases}$$

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:

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

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

$$\tilde{f}(w) \stackrel{\text{def}}{=} \int_0^\infty F(\alpha) d\alpha \tag{16.17}$$

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

$$\tilde{f}(w) \stackrel{\text{def}}{=} \int_0^\infty F(\alpha) d\alpha \tag{16.17}$$

$$= \int_0^\infty f(\{e \in E : w_e > \alpha\}) d\alpha \tag{16.18}$$

The "integral" in the Choquet integral

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

$$\tilde{f}(w) \stackrel{\text{def}}{=} \int_0^\infty F(\alpha) d\alpha \tag{16.17}$$

$$= \int_0^\infty f(\{e \in E : w_e > \alpha\}) d\alpha \tag{16.18}$$

$$= \int_{w_{m+1}}^\infty f(\{e \in E : w_e > \alpha\}) d\alpha \tag{16.19}$$

The "integral" in the Choquet integral

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

$$\tilde{f}(w) \stackrel{\text{def}}{=} \int_0^\infty F(\alpha) d\alpha \tag{16.17}$$

$$= \int_0^\infty f(\{e \in E : w_e > \alpha\}) d\alpha \tag{16.18}$$

$$= \int_{w_{m+1}}^\infty f(\{e \in E : w_e > \alpha\}) d\alpha \tag{16.19}$$

$$= \sum_{i=1}^m \int_{w_{i+1}}^{w_i} f(\{e \in E : w_e > \alpha\}) d\alpha \tag{16.20}$$

The "integral" in the Choquet integral

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

$$\tilde{f}(w) \stackrel{\text{def}}{=} \int_0^\infty F(\alpha) d\alpha \tag{16.17}$$

$$= \int_0^\infty f(\{e \in E : w_e > \alpha\}) d\alpha \tag{16.18}$$

$$= \int_{w_{m+1}}^\infty f(\{e \in E : w_e > \alpha\}) d\alpha \tag{16.19}$$

$$= \sum_{i=1}^m \int_{w_{i+1}}^{w_i} f(\{e \in E : w_e > \alpha\}) d\alpha \tag{16.20}$$

$$= \sum_{i=1}^m \int_{w_{i+1}}^{w_i} f(E_i) d\alpha = \sum_{i=1}^m f(E_i)(w_i - w_{i+1}) \tag{16.21}$$

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 16.4.2

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

$$\tilde{f}(w) \stackrel{\text{def}}{=} \int_0^\infty F(\alpha) d\alpha \tag{16.22}$$

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 16.4.2

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

$$\tilde{f}(w) \stackrel{\text{def}}{=} \int_0^\infty F(\alpha) d\alpha \tag{16.22}$$

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.

The "integral" in the Choquet integral

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

Definition 16.4.2

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

$$\tilde{f}(w) \stackrel{\text{def}}{=} \int_0^\infty F(\alpha) d\alpha \tag{16.22}$$

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.
- The above integral will be further generalized a bit later.

Choquet integral and aggregation

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

$$\mathsf{AG}(x) = \sum_{A \subseteq E} x(A) w_A = \sum_{A \subseteq E} x(A) \mathsf{AG}(\mathbf{1}_A) \tag{16.23}$$

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:

$$\mathsf{AG}(x) = \sum_{A \subseteq E} x(A) w_A = \sum_{A \subseteq E} x(A) \mathsf{AG}(\mathbf{1}_A) \tag{16.23}$$

  how does this correspond to Lovász extension?

- Let us partition the hypercube $[0,1]^m$ into q polytopes, ~~each~~ the i^{th} defined by a set of vertices ~~$\mathcal{V}_1, \mathcal{V}_2, \dots, \mathcal{V}_q$~~. $\mathcal{V}_i$

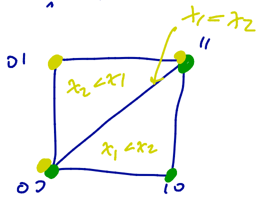
Choquet integral and aggregation

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

$$\mathsf{AG}(x) = \sum_{A \subseteq E} x(A) w_A = \sum_{A \subseteq E} x(A) \mathsf{AG}(\mathbf{1}_A) \tag{16.23}$$

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 $\mathcal{V}_i$ defines the i^{th} polytope.



Choquet integral and aggregation

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

$$\text{AG}(x) = \sum_{A \subseteq E} x(A) w_A = \sum_{A \subseteq E} x(A) \text{AG}(\mathbf{1}_A) \tag{16.23}$$

  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.

Choquet integral and aggregation

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

$$\text{AG}(x) = \sum_{A \subseteq E} x(A) w_A = \sum_{A \subseteq E} x(A)\text{AG}(\mathbf{1}_A) \tag{16.23}$$

  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.
- 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 $\mathbf{1}_A \in \text{conv}(\mathcal{V}(x))$. The affine transformation is as follows:

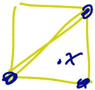
$$\alpha_0^x(A) + \sum_{j=1}^{m} \alpha_j^x(A) x_j \in \mathbb{R} \tag{16.24}$$

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 $\mathbf{1}_A \in \text{conv}(\mathcal{V}(x))$. The affine transformation is as follows:

$$\alpha_0^x(A) + \sum_{j=1}^m \alpha_j^x(A) x_j \in \mathbb{R} \tag{16.24}$$



Note that many of these coefficient are often zero.

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

$$\text{AG}(x) \stackrel{\text{def}}{=} \sum_{A: \mathbf{1}_A \in \mathcal{V}(x)} \left(\alpha_0^x(A) + \sum_{j=1}^m \alpha_j^x(A) x_j \right) \text{AG}(\mathbf{1}_A) \tag{16.25}$$

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

$$\text{conv}(\mathcal{V}_\sigma) = \left\{x \in [0,1]^n | x_{\sigma(1)} \geq x_{\sigma(2)} \geq \cdots \geq x_{\sigma(m)}\right\} \tag{16.26}$$

  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

$$\text{conv}(\mathcal{V}_\sigma) = \left\{x \in [0,1]^n | x_{\sigma(1)} \geq x_{\sigma(2)} \geq \cdots \geq x_{\sigma(m)}\right\} \tag{16.26}$$

  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 $\text{conv}(\mathcal{V}_\sigma)$ 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

$$\text{conv}(\mathcal{V}_\sigma) = \left\{x \in [0,1]^n | x_{\sigma(1)} \geq x_{\sigma(2)} \geq \cdots \geq x_{\sigma(m)}\right\} \tag{16.26}$$

  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 $\text{conv}(\mathcal{V}_\sigma)$ 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

$$\text{conv}(\mathcal{V}_\sigma) = \left\{x \in [0,1]^n | x_{\sigma(1)} \geq x_{\sigma(2)} \geq \cdots \geq x_{\sigma(m)}\right\} \tag{16.26}$$

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

Proposition 16.4.3

The above linear interpolation in Eqn. (16.25) *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.

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

$$\text{conv}(\mathcal{V}_\sigma) = \left\{x \in [0,1]^n | x_{\sigma(1)} \geq x_{\sigma(2)} \geq \cdots \geq x_{\sigma(m)}\right\} \tag{16.26}$$

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

Proposition 16.4.3

The above linear interpolation in Eqn. (16.25) 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.

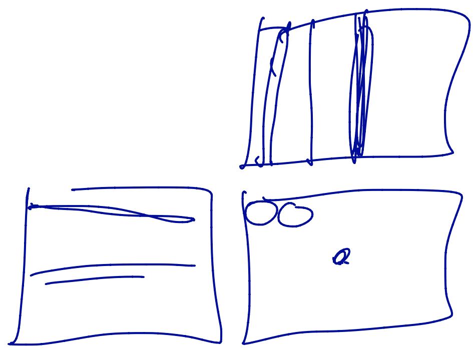
- Hence, Lovász extension is a generalized aggregation function.

$$f(A) = \frac{1}{m} |A|$$

$$\check{f}(x) = \frac{1}{m} \sum_{i=1}^{m} x(i)$$

$$f(A) = m(A)$$

$$\check{f}(x) = \sum_{i=1}^{m} m(i) \cdot x(i)$$



Lovász extension as max over orders

- We can also write the Lovász extension as follows:

$$\tilde{f}(w) = \max_{\sigma \in \Pi_{[m]}} w^\intercal c^\sigma \tag{16.27}$$

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:

$$c_i^\sigma = f(E_{\sigma_i}) - f(E_{\sigma_{i-1}}) \tag{16.28}$$

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:

$$\tilde{f}(w) = \max_{\sigma \in \Pi_{[m]}} w^\intercal c^\sigma \tag{16.27}$$

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:

$$c^\sigma_i = f(E_{\sigma_i}) - f(E_{\sigma_{i-1}}) \tag{16.28}$$

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

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

$$\tilde{f}(w) = \max_{x \in P_f} w^\intercal x = \max_{x \in B_f} w^\intercal x \tag{16.29}$$

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.

$x \in \mathbb{R}$

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:

$$\tilde{f}(w) = \sum_{i=1}^{m} w(e_i) f(e_i | E_{i-1}) \tag{16.30}$$

$$= \sum_{i=1}^{m-1} f(E_i)(w(e_i) - w(e_{i+1})) + f(E) w(e_m) a \tag{16.31}$$

$$= \sum_{i=1}^{m-1} \lambda_i f(E_i) \tag{16.32}$$

Lovász extension, as integral

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

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

$$= \sum_{i=1}^{m-1} f(E_i)(w(e_i) - w(e_{i+1})) + f(E) w(e_m) \tag{16.34}$$

$f(\{e \in E : w(e) \geq \alpha\})$

$$= \int_{\min\{w_1,\dots,w_m\}}^{+\infty} f(\{w \geq \alpha\}) d\alpha + f(E) \min\{w_1, \dots, w_m\} \tag{16.35}$$

$$= \int_{0}^{+\infty} f(\{w \geq \alpha\}) d\alpha + \int_{-\infty}^{0} [f(\{w \geq \alpha\}) - f(E)] d\alpha \tag{16.36}$$

general Lovász extension, as simple integral

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

$$\tilde{f}(w) = \int_{-\infty}^{+\infty} \hat{f}(\alpha) d\alpha \tag{16.37}$$

where

$$\hat{f}(\alpha) = \begin{cases} f(\{w \geq \alpha\}) & \text{if } \alpha >= 0 \\ f(\{w \geq \alpha\}) - f(E) & \text{if } \alpha < 0 \end{cases} \tag{16.38}$$

general Lovász extension, as simple integral

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

$$\tilde{f}(w) = \int_{-\infty}^{+\infty} \hat{f}(\alpha) d\alpha \tag{16.37}$$

where

$$\hat{f}(\alpha) = \begin{cases} f(\{w \geq \alpha\}) & \text{if } \alpha >= 0 \\ f(\{w \geq \alpha\}) - f(E) & \text{if } \alpha < 0 \end{cases} \tag{16.38}$$

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

$$\tilde{f}(w) = \int_{-\infty}^{+\infty} \hat{f}(\alpha) d\alpha \tag{16.37}$$

where

$$\hat{f}(\alpha) = \begin{cases} f(\{w \geq \alpha\}) & \text{if } \alpha >= 0 \\ f(\{w \geq \alpha\}) - f(E) & \text{if } \alpha < 0 \end{cases} \tag{16.38}$$

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

Lovász extension, as integral

- To show Eqn. (16.35), first note that the r.h.s. terms are the same since $w(e_m) = \min\{w_1, \dots, w_m\}$.

Lovász extension, as integral

- To show Eqn. (16.35), 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

$$f(\{w \geq \alpha\}) = \begin{cases} 0 & \text{if } \alpha > w(e_1) \\ f(E_k) & \text{if } \alpha \in (w(e_{k+1}), w(e_k)), k \in \{1, \ldots, m-1\} \\ f(E) & \text{if } \alpha < w(e_m) \end{cases} \tag{16.39}$$

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

Lovász extension, as integral

- To show Eqn. (16.35), first note that the r.h.s. terms are the same since $w(e_m) = \min\{w_1, \dots, w_m\}$.
- Then, consider that, as a function of α, we have

$$f(\{w \geq \alpha\}) = \begin{cases} 0 & \text{if } \alpha > w(e_1) \\ f(E_k) & \text{if } \alpha \in (w(e_{k+1}), w(e_k)), k \in \{1, \dots, m-1\} \\ f(E) & \text{if } \alpha < w(e_m) \end{cases} \tag{16.39}$$

  we may use open intervals since sets of zero measure don't change integration.
- Inside the integral, then, this recovers Eqn. (16.34).

Lovász extension, as integral

- To show Eqn. (16.36), start with Eqn. (16.35), note $w_m = \min\{w_1, \ldots, w_m\}$, take any $\beta \leq \min\{0, w_1, \ldots, w_m\}$, and form:

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

Lovász extension, as integral

- To show Eqn. (16.36), start with Eqn. (16.35), note $w_m = \min\{w_1, \dots, w_m\}$, take any $\beta \leq \min\{0, w_1, \dots, w_m\}$, and form:

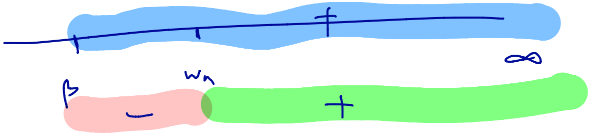
$$\tilde{f}(w) = \int_{w_m}^{+\infty} f(\{w \geq \alpha\}) d\alpha + f(E) \min\{w_1, \dots, w_m\}$$

Lovász extension, as integral

- To show Eqn. (16.36), start with Eqn. (16.35), 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. (16.36), start with Eqn. (16.35), 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

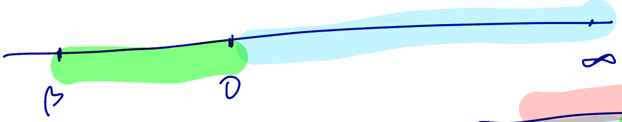
- To show Eqn. (16.36), start with Eqn. (16.35), note $w_m = \min\{w_1, \ldots, w_m\}$, take any $\beta \leq \min\{0, w_1, \ldots, w_m\}$, and form:

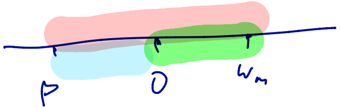
$$\tilde{f}(w) = \int_{w_m}^{+\infty} f(\{w \geq \alpha\})d\alpha + f(E)\min\{w_1, \ldots, w_m\}$$

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

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

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





Lovász extension, as integral

- To show Eqn. (16.36), start with Eqn. (16.35), 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. (16.36), start with Eqn. (16.35), note $w_m = \min\{w_1, \dots, w_m\}$, take any $\beta \leq \min\{0, w_1, \dots, w_m\}$, and form:

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

and then let $\beta \to -\infty$ and we get Eqn. (16.36), i.e.:

Lovász extension, as integral

- To show Eqn. (16.36), start with Eqn. (16.35), note $w_m = \min\{w_1, \ldots, w_m\}$, take any $\beta \le \min\{0, w_1, \ldots, w_m\}$, and form:

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

and then let $\beta \to -\infty$ and we get Eqn. (16.36), 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 16.5.1

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 16.5.1

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 16.5.1

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

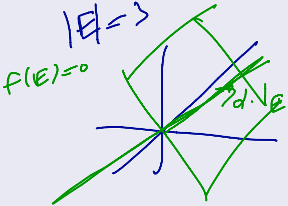
Lovász extension properties

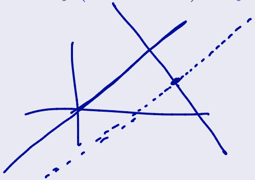
- Using the above, have the following (some of which we've seen):

Theorem 16.5.1

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 16.5.1

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 16.5.1

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 16.5.1

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 16.5.1

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. (16.33) through (16.36), 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

$$\tilde{f}(-w) = \int_{-\infty}^{\infty} f(\{-w \geq \alpha\})d\alpha \tag{16.42}$$

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 properties

- Given Eqns. (16.33) through (16.36), 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

$$\tilde{f}(-w) = \int_{-\infty}^{\infty} f(\{-w \geq \alpha\})d\alpha = \int_{-\infty}^{\infty} f(\{w \leq -\alpha\})d\alpha \quad (16.40)$$

(16.42)

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 properties

- Given Eqns. (16.33) through (16.36), 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

$$\tilde{f}(-w) = \int_{-\infty}^{\infty} f(\{-w \geq \alpha\})d\alpha = \int_{-\infty}^{\infty} f(\{w \leq -\alpha\})d\alpha \quad (16.40)$$

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

$$(16.42)$$

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 properties

- Given Eqns. (16.33) through (16.36), most of the above properties are relatively easy to derive.
- For example, if f is symmetric, and since $f(E) = f(\emptyset) = 0$, we have

$$\tilde{f}(-w) = \int_{-\infty}^{\infty} f(\{-w \geq \alpha\})d\alpha = \int_{-\infty}^{\infty} f(\{w \leq -\alpha\})d\alpha \tag{16.40}$$

$$\stackrel{(a)}{=} \int_{-\infty}^{\infty} f(\{w \leq \alpha\})d\alpha \stackrel{(b)}{=} \int_{-\infty}^{\infty} f(\{w > \alpha\})d\alpha \tag{16.41}$$

(16.42)

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 properties

- Given Eqns. (16.33) through (16.36), 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

$$\tilde{f}(-w) = \int_{-\infty}^{\infty} f(\{-w \geq \alpha\})d\alpha = \int_{-\infty}^{\infty} f(\{w \leq -\alpha\})d\alpha \quad (16.40)$$

$$\stackrel{(a)}{=} \int_{-\infty}^{\infty} f(\{w \leq \alpha\})d\alpha \stackrel{(b)}{=} \int_{-\infty}^{\infty} f(\{w > \alpha\})d\alpha \quad (16.41)$$

$$= \int_{-\infty}^{\infty} f(\{w \geq \alpha\})d\alpha \quad (16.42)$$

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 properties

- Given Eqns. (16.33) through (16.36), most of the above properties are relatively easy to derive.
- For example, if f is symmetric, and since $f(E) = f(\emptyset) = 0$, we have

$$\tilde{f}(-w) = \int_{-\infty}^{\infty} f(\{-w \geq \alpha\})d\alpha = \int_{-\infty}^{\infty} f(\{w \leq -\alpha\})d\alpha \tag{16.40}$$

$$\stackrel{(a)}{=} \int_{-\infty}^{\infty} f(\{w \leq \alpha\})d\alpha \stackrel{(b)}{=} \int_{-\infty}^{\infty} f(\{w > \alpha\})d\alpha \tag{16.41}$$

$$= \int_{-\infty}^{\infty} f(\{w \geq \alpha\})d\alpha = \tilde{f}(w) \tag{16.42}$$

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 $\mathbb{E}[h(\alpha)] = \int_0^1 h(\alpha)d\alpha$.

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 $\mathbb{E}[h(\alpha)] = \int_0^1 h(\alpha) d\alpha$.
- Hence, for $w \in [0,1]^m$, we can also define the Lovász extension as

$$\tilde{f}(w) = \mathbb{E}_{p(\alpha)}[\underbrace{f(\{w \geq \alpha\})}_{h(\alpha)}] = \mathbb{E}_{p(\alpha)}[\underbrace{f(e \in E : w(e_i) \geq \alpha)}_{h(\alpha)}] \quad (16.43)$$

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

$$\tilde{f}(w) = \mathbb{E}_{p(\alpha)}[\underbrace{f(\{w \geq \alpha\})}_{h(\alpha)}] = \mathbb{E}_{p(\alpha)}[\underbrace{f(e \in E : w(e_i) \geq \alpha)}_{h(\alpha)}] \quad (16.43)$$

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

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

- If $w_1 \geq w_2$, then

$$\tilde{f}(w) = w_1 f(\{1\}) + w_2 f(\{2\}|\{1\}) \quad (16.44)$$
$$= (w_1 - w_2) f(\{1\}) + w_2 f(\{1, 2\}) \quad (16.45)$$

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

- If $w_1 \geq w_2$, then

$$\tilde{f}(w) = w_1 f(\{1\}) + w_2 f(\{2\}|\{1\}) \quad (16.44)$$

$$= (w_1 - w_2) f(\{1\}) + w_2 f(\{1, 2\}) \quad (16.45)$$

- If $w_1 \leq w_2$, then

$$\tilde{f}(w) = w_2 f(\{2\}) + w_1 f(\{1\}|\{2\}) \quad (16.46)$$

$$= (w_2 - w_1) f(\{2\}) + w_1 f(\{1, 2\}) \quad (16.47)$$

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

- If $w_1 \geq w_2$, then

$$\tilde{f}(w) = w_1 f(\{1\}) + w_2 f(\{2\}|\{1\}) \tag{16.48}$$

$$= (w_1 - w_2) f(\{1\}) + w_2 f(\{1, 2\}) \tag{16.49}$$

$$= \frac{1}{2} f(1)(w_1 - w_2) + \frac{1}{2} f(1)(w_1 - w_2) \tag{16.50}$$

$$+ \frac{1}{2} f(\{1, 2\})(w_1 + w_2) - \frac{1}{2} f(\{1, 2\})(w_1 - w_2) \tag{16.51}$$

$$+ \frac{1}{2} f(2)(w_1 - w_2) + \frac{1}{2} f(2)(w_2 - w_1) \tag{16.52}$$

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

- If $w_1 \geq w_2$, then

$$\begin{aligned}
\tilde{f}(w) &= w_1 f(\{1\}) + w_2 f(\{2\}|\{1\}) && (16.48)\\
&= (w_1 - w_2) f(\{1\}) + w_2 f(\{1,2\}) && (16.49)\\
&= \frac{1}{2} f(1)(w_1 - w_2) + \frac{1}{2} f(1)(w_1 - w_2) && (16.50)\\
&\quad + \frac{1}{2} f(\{1,2\})(w_1 + w_2) - \frac{1}{2} f(\{1,2\})(w_1 - w_2) && (16.51)\\
&\quad + \frac{1}{2} f(2)(w_1 - w_2) + \frac{1}{2} f(2)(w_2 - w_1) && (16.52)
\end{aligned}$$

- 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

$$\tilde{f}(w) = \frac{1}{2}\left(f(\{1\}) + f(\{2\}) - f(\{1,2\})\right)|w_1 - w_2| \tag{16.53}$$

$$+ \frac{1}{2}\left(f(\{1\}) - f(\{2\}) + f(\{1,2\})\right) w_1 \tag{16.54}$$

$$+ \frac{1}{2}\left(-f(\{1\}) + f(\{2\}) + f(\{1,2\})\right) w_2 \tag{16.55}$$

$$= -\left(f(\{1\}) + f(\{2\}) - f(\{1,2\})\right) \min\{w_1, w_2\} \tag{16.56}$$

$$+ f(\{1\})w_1 + f(\{2\})w_2 \tag{16.57}$$

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

- This gives, for general w_1, w_2, that

$$\tilde{f}(w) = \frac{1}{2}\left(f(\{1\}) + f(\{2\}) - f(\{1,2\})\right)|w_1 - w_2| \quad (16.53)$$

$$+ \frac{1}{2}\left(f(\{1\}) - f(\{2\}) + f(\{1,2\})\right)w_1 \quad (16.54)$$

$$+ \frac{1}{2}\left(-f(\{1\}) + f(\{2\}) + f(\{1,2\})\right)w_2 \quad (16.55)$$

$$= -\left(f(\{1\}) + f(\{2\}) - f(\{1,2\})\right)\min\{w_1, w_2\} \quad (16.56)$$

$$+ f(\{1\})w_1 + f(\{2\})w_2 \quad (16.57)$$

- 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

$$\tilde{f}(w) = \frac{1}{2}\left(f(\{1\}) + f(\{2\}) - f(\{1,2\})\right)|w_1 - w_2| \quad (16.53)$$

$$+ \frac{1}{2}\left(f(\{1\}) - f(\{2\}) + f(\{1,2\})\right) w_1 \quad (16.54)$$

$$+ \frac{1}{2}\left(-f(\{1\}) + f(\{2\}) + f(\{1,2\})\right) w_2 \quad (16.55)$$

$$= -\left(f(\{1\}) + f(\{2\}) - f(\{1,2\})\right) \min\{w_1, w_2\} \quad (16.56)$$

$$+ f(\{1\})w_1 + f(\{2\})w_2 \quad (16.57)$$

- 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

$$\tilde{f}(w) = w_1 f(\{1\}) + w_2 f(\{2\}|\{1\}) \tag{16.58}$$

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

- If $w_1 \geq w_2$, then

$$\tilde{f}(w) = w_1 f(\{1\}) + w_2 f(\{2\}|\{1\}) \tag{16.58}$$

  - 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

$$\tilde{f}(w) = w_1 f(\{1\}) + w_2 f(\{2\}|\{1\}) \tag{16.58}$$

  - 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

$$\tilde{f}(w) = w_1 f(\{1\}) + w_2 f(\{2\}|\{1\}) \tag{16.58}$$

  - 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

$$\tilde{f}(w) = w_2 f(\{2\}) + w_1 f(\{1\}|\{2\}) \tag{16.59}$$

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

- If $w_1 \geq w_2$, then

$$\tilde{f}(w) = w_1 f(\{1\}) + w_2 f(\{2\}|\{1\}) \tag{16.58}$$

  - 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

$$\tilde{f}(w) = w_2 f(\{2\}) + w_1 f(\{1\}|\{2\}) \tag{16.59}$$

  - 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

$$\tilde{f}(w) = w_1 f(\{1\}) + w_2 f(\{2\}|\{1\}) \tag{16.58}$$

  - 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

$$\tilde{f}(w) = w_2 f(\{2\}) + w_1 f(\{1\}|\{2\}) \tag{16.59}$$

  - 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

$$\tilde{f}(w) = w_1 f(\{1\}) + w_2 f(\{2\}|\{1\}) \tag{16.58}$$

  - 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

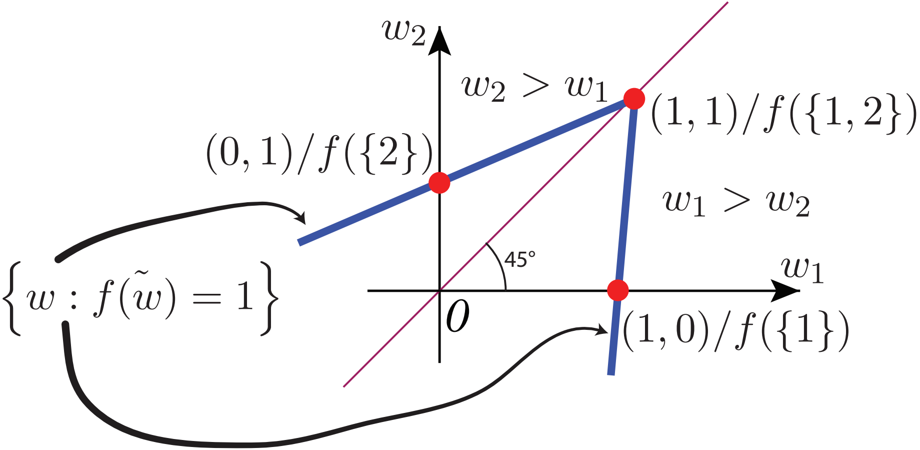
$$\tilde{f}(w) = w_2 f(\{2\}) + w_1 f(\{1\}|\{2\}) \tag{16.59}$$

  - 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 $\tilde{f}(w + \alpha \mathbf{1}_E) = \tilde{f}(w)$ when $f(E) = 0$.

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

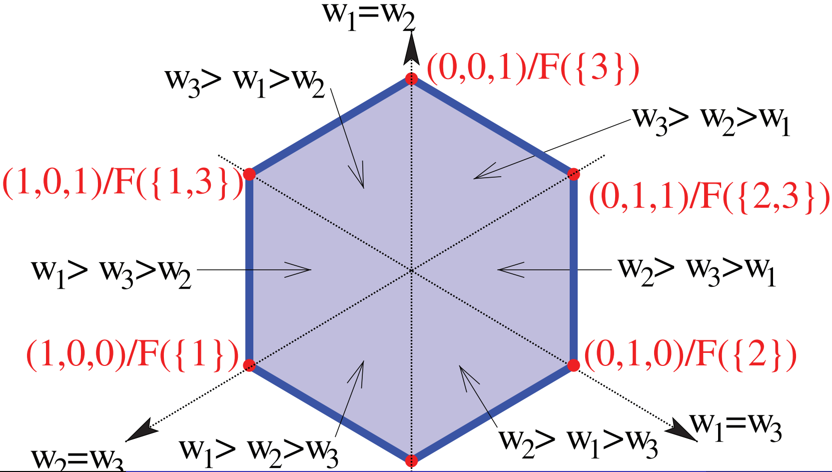
- In order to visualize in 3D, we make a few simplifications.
- Consider any submodular f' and $x \in B_{f'}$. Then $f(A) = f'(A) - x(A)$ is submodular, and moreover $f(E) = f'(E) - x(E) = 0$.
- Hence, from $\tilde{f}(w + \alpha \mathbf{1}_E) = \tilde{f}(w) + \alpha f(E)$, we have that $\tilde{f}(w + \alpha \mathbf{1}_E) = \tilde{f}(w)$ when $f(E) = 0$.
- 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 $\tilde{f}(w) = \max_{k \in \{1,2,3\}} w_k - \min_{k \in \{1,2,3\}} w_k$.

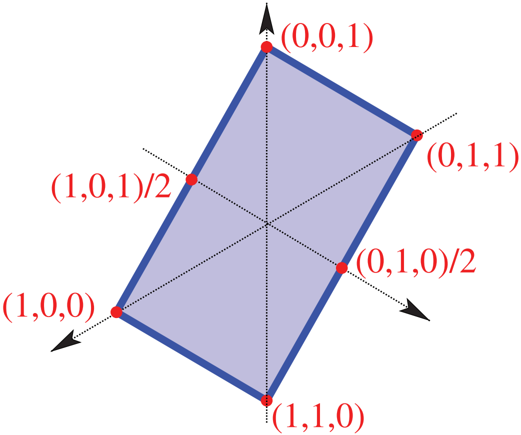
Example: $m = 3$, $E = \{1, 2, 3\}$

- 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 $\tilde{f}(w) = \max_{k \in \{1,2,3\}} w_k - \min_{k \in \{1,2,3\}} w_k$.



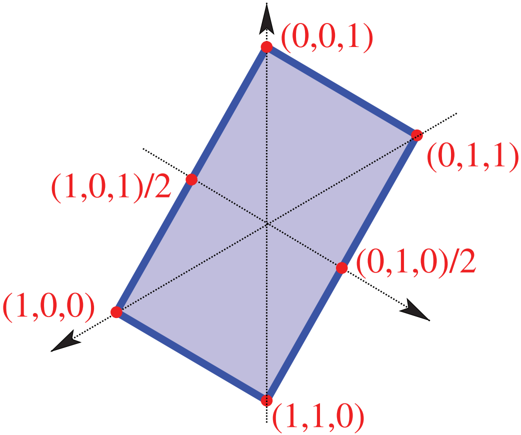
Example: $m = 3$, $E = \{1, 2, 3\}$

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



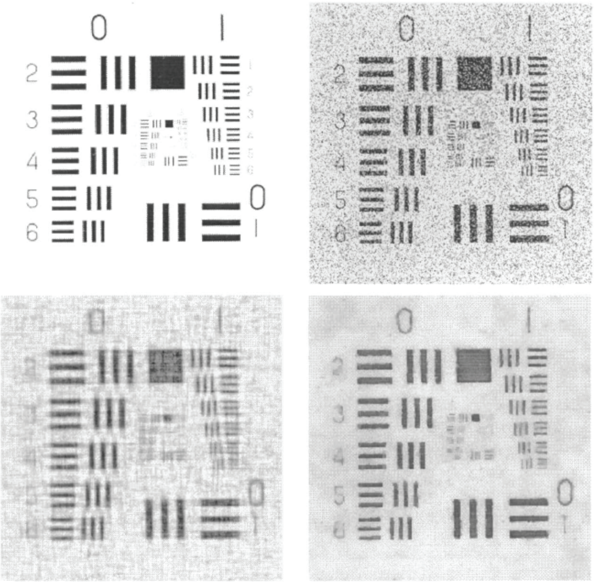
Example: $m = 3$, $E = \{1, 2, 3\}$

- 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.



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

$$\tilde{f}(w) = \sum_{i=1}^{m} w(e_i)\big(g(M_i) - g(M_{i-1})\big) \tag{16.60}$$

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

$$\tilde{f}(w) = \sum_{i=1}^{m} w(e_i)\big(g(M_i) - g(M_{i-1})\big) \tag{16.60}$$

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

$$\tilde{f}(w) = \sum_{i=1}^{m} w(e_i)\big(g(i) - g(i-1)\big) \tag{16.61}$$

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 $m_{ij} = m((i, j))$:

$$f(X) = \sum_{i \in X, j \in V \setminus X} m_{ij} \tag{16.62}$$

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 $m_{ij} = m((i,j))$:

$$f(X) = \sum_{i \in X, j \in V \setminus X} m_{ij} \tag{16.62}$$

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

$$\tilde{f}(w) = \sum_{i,j \in V} m_{ij} \max\{(w_i - w_j), 0\} \tag{16.63}$$

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 $m_{ij} = m((i, j))$:

$$f(X) = \sum_{i \in X, j \in V \setminus X} m_{ij} \quad (16.62)$$

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

$$\tilde{f}(w) = \sum_{i,j \in V} m_{ij} \max \{(w_i - w_j), 0\} \quad (16.63)$$

  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:

$$\min_{w \in \mathbb{R}^n} \frac{1}{m} \sum_{i=1}^m \ell(y_i, w^\intercal x_i) + \lambda \Omega(w), \tag{16.64}$$

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:

$$\min_{w^1, \dots, w^k \in \mathbb{R}^n} \sum_{j=1}^k \frac{1}{m} \sum_{i=1}^m \ell(y_i^k, (w^k)^\intercal x_i) + \lambda \Omega(w^k), \tag{16.65}$$

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

$$\min_{x_1, \dots, x_m} \min_{w^1, \dots, w^k \in \mathbb{R}^n} \sum_{j=1}^k \frac{1}{m} \sum_{i=1}^m \ell(y_i^k, (w^k)^\intercal x_i) + \lambda \Omega(w^k), \tag{16.66}$$

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:

$$\Omega(w) = \sum_{i=2}^{N} |w_i - w_{i-1}| \tag{16.67}$$

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:

$$\Omega(w) = \sum_{i=2}^{N} |w_i - w_{i-1}| \tag{16.67}$$

- 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, $\|w\|_0 = \mathbf{1}^\intercal \text{supp}(w)$.

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, $\|w\|_0 = \mathbf{1}^\intercal \text{supp}(w)$.
- Using $\Omega(w) = \|w\|_0$ is NP-hard, instead we often optimize tightest convex relaxation, $\|w\|_1$ 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, $\|w\|_0 = \mathbf{1}^\intercal \text{supp}(w)$.
- Using $\Omega(w) = \|w\|_0$ is NP-hard, instead we often optimize tightest convex relaxation, $\|w\|_1$ 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, $\|w\|_0 = \mathbf{1}^\intercal \text{supp}(w)$.
- Using $\Omega(w) = \|w\|_0$ is NP-hard, instead we often optimize tightest convex relaxation, $\|w\|_1$ 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, $\|w\|_0 = \mathbf{1}^\intercal \text{supp}(w)$.
- Using $\Omega(w) = \|w\|_0$ is NP-hard, instead we often optimize tightest convex relaxation, $\|w\|_1$ 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 (Vondrák 2007, 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, $\|w\|_0 = \mathbf{1}^\intercal \text{supp}(w)$.
- Using $\Omega(w) = \|w\|_0$ is NP-hard, instead we often optimize tightest convex relaxation, $\|w\|_1$ 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 (Vondrák 2007, 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, $\|w\|_0 = \mathbf{1}^\intercal \text{supp}(w)$.
- Using $\Omega(w) = \|w\|_0$ is NP-hard, instead we often optimize tightest convex relaxation, $\|w\|_1$ 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 (Vondrák 2007, 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 function 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 function 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 function 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 function 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 function 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 function 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.

Concave closure

- The concave closure is defined as:

$$\hat{f}(x) = \max_{p \in \triangle^n(x)} \sum_{S \subseteq V} p_S f(S) \tag{16.68}$$

where $\triangle^n(x) = \left\{ p \in \mathbb{R}^{2^n} : \sum_{S \subseteq V} p_S = 1, p_S \geq 0 \forall S \subseteq V, \; \& \; \sum_{S \subseteq V} p_S \mathbf{1}_S = x \right\}$

Concave closure

- The concave closure is defined as:

$$\hat{f}(x) = \max_{p \in \triangle^n(x)} \sum_{S \subseteq V} p_S f(S) \tag{16.68}$$

where $\triangle^n(x) =$
$\left\{ p \in \mathbb{R}^{2^n} : \sum_{S \subseteq V} p_S = 1, p_S \geq 0 \forall S \subseteq V, \ \& \ \sum_{S \subseteq V} p_S \mathbf{1}_S = x \right\}$

- This is tight at the hypercube vertices, concave, and the concave envelope for the dual reasons as the convex closure.

Concave closure

- The concave closure is defined as:

$$\hat{f}(x) = \max_{p \in \triangle^n(x)} \sum_{S \subseteq V} p_S f(S) \tag{16.68}$$

where $\triangle^n(x) =$
$\left\{ p \in \mathbb{R}^{2^n} : \sum_{S \subseteq V} p_S = 1, p_S \geq 0 \forall S \subseteq V, \text{ \& } \sum_{S \subseteq V} p_S \mathbf{1}_S = x \right\}$

- This is tight at the hypercube vertices, concave, and the concave envolope for the dual reasons as the convex closure.
- Unlike the convex extension, the concave closure is defined by the Lovász extension iff f is a supermodular function.

Concave closure

- The concave closure is defined as:

$$\hat{f}(x) = \max_{p \in \triangle^n(x)} \sum_{S \subseteq V} p_S f(S) \tag{16.68}$$

where $\triangle^n(x) = \left\{ p \in \mathbb{R}^{2^n} : \sum_{S \subseteq V} p_S = 1, p_S \geq 0 \forall S \subseteq V, \ \& \ \sum_{S \subseteq V} p_S \mathbf{1}_S = x \right\}$

- This is tight at the hypercube vertices, concave, and the concave envolope for the dual reasons as the convex closure.
- Unlike the convex extension, the concave closure is defined by the Lovász extension iff f is a supermodular function.
- When f is submodular, even evaluating $\hat{f}$ is NP-hard (rough intuition: submodular maxmization is NP-hard (reduction to set cover), if we could evaluate $\hat{f}$ in poly time, we can maximize concave function to solve submodular maximization in poly time).

Multilinear extension

- Rather than the concave closure, multi-linear extension is used as a surrogate. For $x \in [0,1]^V = [0,1]^{[n]}$

$$\tilde{f}(x) = \sum_{S \subseteq V} f(S) \prod_{i \in S} x_i \prod_{i \in V \setminus S} (1 - x_i) = E_{S \sim x}[f(S)] \tag{16.69}$$

Multilinear extension

- Rather than the concave closure, multi-linear extension is used as a surrogate. For $x \in [0,1]^V = [0,1]^{[n]}$

$$\tilde{f}(x) = \sum_{S \subseteq V} f(S) \prod_{i \in S} x_i \prod_{i \in V \setminus S} (1 - x_i) = E_{S \sim x}[f(S)] \tag{16.69}$$

- Can be viewed as expected value of $f(S)$ where S is a random set distributed via x, so $\Pr(v \in S) = x_v$ and is independent of $\Pr(u \in S) = x_u$, $v \neq u$.

Multilinear extension

- Rather than the concave closure, multi-linear extension is used as a surrogate. For $x \in [0,1]^V = [0,1]^{[n]}$

$$\tilde{f}(x) = \sum_{S \subseteq V} f(S) \prod_{i \in S} x_i \prod_{i \in V \setminus S} (1 - x_i) = E_{S \sim x}[f(S)] \tag{16.69}$$

- Can be viewed as expected value of $f(S)$ where S is a random set distributed via x, so $\Pr(v \in S) = x_v$ and is independent of $\Pr(u \in S) = x_u$, $v \neq u$.
- This is tight at the hypercube vertices (immediate, since $f(\mathbf{1}_A)$ yields only one term in the sum non-zero, namely the one where $S = A$).

Multilinear extension

- Rather than the concave closure, multi-linear extension is used as a surrogate. For $x \in [0,1]^V = [0,1]^{[n]}$

$$\tilde{f}(x) = \sum_{S \subseteq V} f(S) \prod_{i \in S} x_i \prod_{i \in V \setminus S} (1 - x_i) = E_{S \sim x}[f(S)] \tag{16.69}$$

- Can be viewed as expected value of $f(S)$ where S is a random set distributed via x, so $\Pr(v \in S) = x_v$ and is independent of $\Pr(u \in S) = x_u$, $v \neq u$.
- This is tight at the hypercube vertices (immediate, since $f(\mathbf{1}_A)$ yields only one term in the sum non-zero, namely the one where $S = A$).
- Why called multilinear (multi-linear) extension? It is linear in each of its arguments (i.e., $\tilde{f}(x_1, x_2, \dots, \alpha x_k + \beta x'_k, \dots, x_n) = \alpha \tilde{f}(x_1, x_2, \dots, x_k, \dots, x_n) + \beta \tilde{f}(x_1, x_2, \dots, x'_k, \dots, x_n)$

Multilinear extension

- Rather than the concave closure, multi-linear extension is used as a surrogate. For $x \in [0,1]^V = [0,1]^{[n]}$

$$\tilde{f}(x) = \sum_{S \subseteq V} f(S) \prod_{i \in S} x_i \prod_{i \in V \setminus S} (1 - x_i) = E_{S \sim x}[f(S)] \tag{16.69}$$

- Can be viewed as expected value of $f(S)$ where S is a random set distributed via x, so $\Pr(v \in S) = x_v$ and is independent of $\Pr(u \in S) = x_u$, $v \neq u$.
- This is tight at the hypercube vertices (immediate, since $f(\mathbf{1}_A)$ yields only one term in the sum non-zero, namely the one where $S = A$).
- Why called multilinear (multi-linear) extension? It is linear in each of its arguments (i.e., $\tilde{f}(x_1, x_2, \ldots, \alpha x_k + \beta x'_k, \ldots, x_n) = \alpha \tilde{f}(x_1, x_2, \ldots, x_k, \ldots, x_n) + \beta \tilde{f}(x_1, x_2, \ldots, x'_k, \ldots, x_n)$
- This is unfortunately not concave. However there are some useful properties.

Multilinear extension

Lemma 16.7.1

Let $\tilde{f}(x)$ be the multilinear extension of a set function $f : 2^V \to \mathbb{R}$. Then:

- *If f is monotone non-decreasing, then $\frac{\partial \tilde{f}}{\partial x_v} \geq 0$ for all $v \in V$ within $[0,1]^V$ (i.e., $\tilde{f}$ is also monotone non-decreasing).*
- *If f is submodular, then $\tilde{f}$ has an antitone supergradient, i.e., $\frac{\partial^2 \tilde{f}}{\partial x_i \partial x_j} \leq 0$ for all $i, j \in V$ within $[0,1]^V$.*

Proof.

- First part (monotonicity). Choose $x \in [0,1]^V$ and let $S \sim x$ be random where x is treated as a distribution (so elements v is chosen with probability x_v independently of any other element).

. . .

Multilinear extension

... proof continued.

- Since $\tilde{f}$ is multilinear, derivative is a simple difference when only one argument varies, i.e.,

$$\frac{\partial \tilde{f}}{\partial x_v} = \tilde{f}(x_1, x_2, \dots, x_{v_1}, 1, x_{v+1}, \dots, x_n) \tag{16.70}$$

$$- \tilde{f}(x_1, x_2, \dots, x_{v_1}, 0, x_{v+1}, \dots, x_n) \tag{16.71}$$

$$= E_{S\sim x}[f(S + v)] - E_{S\sim x}[f(S - v)] \tag{16.72}$$

$$\geq 0 \tag{16.73}$$

where the final part follows due to monotonicity of each argument, i.e., $f(S + i) \geq f(S - i)$ for any S and $i \in V$.

□

Multilinear extension

. . . proof continued.

- Second part of proof (antitone supergradient) also relies on simple consequence of multilinearity, namely multilinearity of the derivative as well. In this case

$$\frac{\partial^2 \tilde{f}}{\partial x_i \partial x_j} = \frac{\partial \tilde{f}}{\partial x_j}(x_1, \dots, x_{i-1}, 1, x_{i+1}, \dots, x_n) \tag{16.74}$$

$$- \frac{\partial \tilde{f}}{\partial x_j}(x_1, \dots, x_{i-1}, 0, x_{i+1}, \dots, x_n) \tag{16.75}$$

$$= E_{S \sim x}[f(S + i + j) - f(S + i - j)] \tag{16.76}$$

$$- E_{S \sim x}[f(S - i + j) - f(S - i - j)] \tag{16.77}$$

$$\leq 0 \tag{16.78}$$

since by submodularity, we have

$$f(S + i - j) + f(S - i + j) \geq f(S + i + j) + f(S - i - j) \tag{16.79}$$

□

Multilinear extension: some properties

Corollary 16.7.2

let f be a function and $\tilde{f}$ its multilinear extension on $[0,1]^V$.

- *if f is monotone non-decreasing then $\tilde{f}$ is non-decreasing along any strictly non-negative direction (i.e., $\tilde{f}(x) \leq \tilde{f}(y)$ whenever $x \leq y$, or $\tilde{f}(x) \leq \tilde{f}(x + \epsilon \mathbf{1}_v)$ for any $v \in V$ and any $\epsilon \geq 0$.*
- *If f is submodular, then $\tilde{f}$ is concave along any non-negative direction (i.e., the function $g(\alpha) = \tilde{f}(x + \alpha z)$ is 1-D concave in α for any $z \in \mathbb{R}_+$).*
- *If f is submodular than $\tilde{f}$ is convex along any diagonal direction (i.e., the function $g(\alpha) = \tilde{f}(x + \alpha(\mathbf{1}_v - \mathbf{1}_u))$ is 1-D convex in α for any $u \neq v$.*