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) + f(B) ≥ f(A ∪ B) + f(A ∩ B)

- f(A) + 2f(C) + f(B) - f(A) + f(C) + f(B) - f(A ∩ B)
Announcements, Assignments, and Reminders

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,
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
L16(5/16): More Lovasz extension, Choquet, defs/props, examples, multilinear extension
L17(5/21):
L18(5/23):
L19(5/30):

Last day of instruction, June 1st. Finals Week: June 2-8, 2018.
Convex Closure of Discrete Set Functions

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

\[
\tilde{f}(x) = \min_{p \in \Delta^n(x)} \sum_{S \subseteq V} p_S f(S)
\]

(16.1)

where \( \Delta^n(x) = \{ p \in \mathbb{R}^{2^n} : \sum_{S \subseteq V} p_S = 1, p_S \geq 0 \forall S \subseteq V, \sum_{S \subseteq V} p_S 1_S = x \} \)

- Hence, \( \Delta^n(x) \) 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 \( p \in \Delta^n(x) \), \( E_{S \sim p}(1_S) = \sum_{S \subseteq V} p_S 1_S = x \).

- Hence, \( \tilde{f}(x) = \min_{p \in \Delta^n(x)} E_{S \sim p}[f(S)] \)

- Note, this is not (necessarily) the Lovász extension, rather this is a convex extension.
Convex Closure of Discrete Set Functions

Given, \( \tilde{f}(x) = \min_{p \in \Delta^n(x)} ES \sim p[f(S)] \), we can show:

1. That \( \tilde{f} \) is tight (i.e., \( \forall S \subseteq V \), we have \( \tilde{f}(1_S) = f(S) \)).
2. That \( \tilde{f} \) is convex (and consequently, that any arbitrary set function has a tight convex extension).
3. That the convex closure \( \tilde{f} \) is the convex envelope of the function defined only on the hypercube vertices, and that takes value \( f(S) \) at \( 1_S \).
4. The definition of the Lovász extension of a set function, and that \( \tilde{f} \) 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

$$\tilde{f}(w) = \max(wx : x \in B_f)$$

(16.12)

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

(16.13)

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

(16.14)

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

(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 $w$.  

Prof. Jeff Bilmes
A continuous extension of submodular $f$

- Definition of the continuous extension, once again, for reference:
  \[ \tilde{f}(w) = \max(wx : x \in B_f) \]  
  \[ (16.12) \]

- Therefore, if $f$ is a submodular function, we can write
  \[ \tilde{f}(w) = w(e_m)f(E_m) + \sum_{i=1}^{m-1} (w(e_i) - w(e_{i+1}))f(E_i) \]  
  \[ (16.13) \]
  \[ = \sum_{i=1}^{m} \lambda_i f(E_i) \]  
  \[ (16.14) \]

  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 $\Rightarrow \tilde{f}(w) = \max(wx : x \in P)$ 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 $\tilde{f}(1_A) = f(A), \forall A$, in this way where

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

(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

$$\lambda_i = \begin{cases} w(e_i) - w(e_{i+1}) & \text{if } i < m \\ w(e_m) & \text{if } i = m \end{cases}$$

(16.22)

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

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

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

- This extension is called the Lovász extension!
Summary: comparison of the two extension forms

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

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

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

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

  \[
  \tilde{f}(w) = \sum_{i=1}^{m} \lambda_i f(E_i) \tag{16.26}
  \]

  where \( w = \sum_{i=1}^{m} \lambda_i 1_{E_i} \) 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. (16.25) and Eq. (16.26), we have \( \tilde{f}(1_A) = f(A), \forall A \), but Eq. (16.26), might not be convex.

- Submodularity is sufficient for convexity, but is it necessary?
Theorem 16.2.5

A function $f : 2^E \to \mathbb{R}$ is submodular iff its Lovász extension $\tilde{f}$ of $f$ is convex.

Proof.

- We’ve already seen that if $f$ is submodular, its extension can be written via Eqn.(??) due to the greedy algorithm, and therefore is also equivalent to $\tilde{f}(w) = \max \{wx : x \in P_f\}$, and thus is convex.

- Conversely, suppose the Lovász extension $\tilde{f}(w) = \sum_i \lambda_i f(E_i)$ of some function $f : 2^E \to \mathbb{R}$ is a convex function.

- We note that, based on the extension definition, in particular the definition of the $\{\lambda_i\}_i$, we have that $\tilde{f}(\alpha w) = \alpha \tilde{f}(w)$ for any $\alpha \in \mathbb{R}_+$. I.e., $f$ is a positively homogeneous convex function.

...
Theorem 16.2.5

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

Proof.

- Assume \( f \) is submodular.
- Given \( x \), let \( p^x \) be an achieving argmin in \( \hat{f}(x) \) 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:
  \[
  \begin{align*}
  \bar{p}^x_A &\leftarrow p^x_A - p^x_B \\
  \bar{p}^x_B &\leftarrow p^x_B - p^x_B \\
  \bar{p}^x_{A \cup B} &\leftarrow p^x_{A \cup B} + p^x_B \\
  \bar{p}^x_{A \cap B} &\leftarrow p^x_{A \cap B} + p^x_B
  \end{align*}
  \tag{16.34}
  \]
  \[
  \begin{align*}
  \bar{p}^x_A \not\subseteq B \\
  \bar{p}^x_B \not\subseteq A \\
  \bar{p}^x_{A \cup B} \not\subseteq A \\
  \bar{p}^x_{A \cup B} \not\subseteq B
  \end{align*}
  \tag{16.35}
  \]
  and by submodularity, this does not increase \( \sum_S p^x_S f(S) \).
Next, assume $f$ is not submodular. We must show that the Lovász extension $\tilde{f}(x)$ and the concave closure $\hat{f}(x)$ need not coincide.
Next, assume $f$ is not submodular. We must show that the Lovász extension $\tilde{f}(x)$ and the concave closure $\hat{f}(x)$ need not coincide.

Since $f$ is not submodular, $\exists S$ and $i, j /\in S$ such that

$$f(S) + f(S + i + j) > f(S + i) + f(S + j),$$

a strict violation of submodularity.
Next, assume $f$ is not submodular. We must show that the Lovász extension $\tilde{f}(x)$ and the concave closure $\alpha f(x)$ 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 $x = 1_S + \frac{1}{2} 1_{\{i,j\}}$. 
Next, assume $f$ is not submodular. We must show that the Lovász extension $\tilde{f}(x)$ 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 $x = 1_S + \frac{1}{2}1_{\{i,j\}}$.

Then L.E. has $\tilde{f}(x) = \frac{1}{2}f(S) + \frac{1}{2}f(S + i + j)$ and this $p^x$ is feasible for $\tilde{f}(x)$ with $p_S^x = 1/2$ and $p_{S+i+j}^x = 1/2$.

$p^x \in \Delta(x)$
... proof cont.

Next, assume $f$ is not submodular. We must show that the Lovász extension $\tilde{f}(x)$ 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 $x = 1_S + \frac{1}{2}1_{\{i,j\}}$.

Then L.E. has $\tilde{f}(x) = \frac{1}{2}f(S) + \frac{1}{2}f(S + i + j)$ and this $p^x$ is feasible for $\tilde{f}(x)$ with $p^x_S = 1/2$ and $p^x_{S+i+j} = 1/2$. $\check{f}(x) \in \mathcal{D}(x)$

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

$\exists \mathcal{A} \in A \forall A = x$ $p^x_A \cdot 1_A = x$
Next, assume \( f \) is not submodular. We must show that the Lovász extension \( \tilde{f}(x) \) and the concave closure \( \tilde{\tilde{f}}(x) \) 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 \( x = 1_S + \frac{1}{2} 1\{i,j\} \).

Then L.E. has \( \tilde{f}(x) = \frac{1}{2} f(S) + \frac{1}{2} f(S + i + j) \) and this \( p^x \) is feasible for \( \tilde{f}(x) \) with \( p^x_S = 1/2 \) and \( p^x_{S+i+j} = 1/2 \).

An alternate feasible distribution for \( \tilde{f}(x) \) in the convex closure is
\[
\tilde{p}^x_{S+i} = \tilde{p}^x_{S+j} = 1/2.
\]

This gives
\[
\tilde{f}(x) \leq \frac{1}{2} [f(S + i) + f(S + j)] < \tilde{\tilde{f}}(x)
\]
meaning \( \tilde{f}(x) \neq \tilde{\tilde{f}}(x) \).

\[
f(x) = x + e^{-x}
\]
Integration and Aggregation

- Integration is just summation (e.g., the ∫ symbol has as its origins a sum).
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 \( \mu \) of sets. E.g., given measurable function \( f \), we can define

\[
\int_X f \, du = \sup I_X(s)
\]

(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:

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

$$\text{WAVG}(x) = \sum_{e \in E} x(e)w(e)$$  \hspace{1cm} (16.3)

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

$$\text{WAVG}(1_e) = w(e)$$  \hspace{1cm} (16.4)
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:

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

Consider $1_e$ for $e \in E$, we have

$$\text{WAVG}(1_e) = w(e)$$  \hfill (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., $\{1_e : e \in E\}$. 

$|E| = 3$
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:

\[
\text{WAVG}(x) = \sum_{e \in E} x(e)w(e)
\]  

(16.3)

Consider \( 1_e \) for \( e \in E \), we have

\[
\text{WAVG}(1_e) = w(e)
\]  

(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., \( \{1_e : e \in E\} \). Moreover, we are interpolating as in

\[
\text{WAVG}(x) = \sum_{e \in E} x(e)w(e) = \sum_{e \in E} x(e)\text{WAVG}(1_e)
\]  

(16.5)
Integration, Aggregation, and Weighted Averages

\[
WAVG(x) = \sum_{e \in E} x(e)w(e) \quad (16.6)
\]

- 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 \(\alpha \in \mathbb{R}\),

\[
WAVG_{w_1 + w_2}(x) = WAVG_{w_1}(x) + WAVG_{w_2}(x), \quad (16.7)
\]

\[
WAVG_w(x_1 + x_2) = WAVG_w(x_1) + WAVG_w(x_2), \quad (16.8)
\]

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

\[
WAVG(\alpha x) = \alpha WAVG(x). \quad (16.9)
\]
Integration, Aggregation, and Weighted Averages

\[ f'(x) = \sum \alpha_i f(e_i) \]

\[ \text{WAVG}(x) = \sum_{e \in E} x(e) w(e) \quad (16.6) \]

- 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 \( \alpha \in \mathbb{R} \),

\[ \text{WAVG}_{w_1 + w_2}(x) = \text{WAVG}_{w_1}(x) + \text{WAVG}_{w_2}(x), \quad (16.7) \]
\[ \text{WAVG}_w(x_1 + x_2) = \text{WAVG}_w(x_1) + \text{WAVG}_w(x_2), \quad (16.8) \]

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

\[ \text{WAVG}(\alpha x) = \alpha \text{WAVG}(x). \]

- How related? The Lovász extension \( f'(x) \) 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 \( \alpha \geq 0 \)).
More complex “nonlinear” aggregation functions can be constructed by defining the aggregation function on all vertices of the hypercube. I.e., for each $1_A : A \subseteq E$ we might have (for all $A \subseteq E$):

$$AG(1_A) = w_A$$  \hspace{1cm} (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 \( 1_A : A \subseteq E \) we might have (for all \( A \subseteq E \)):

\[
AG(1_A) = w_A \in \mathbb{R} \quad \left( \sum_{A \subseteq E} w_A \right) \geq 2 \quad (16.10)
\]

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

\[
AG(x) = \sum_{A \subseteq E} x(A)w_A = \sum_{A \subseteq E} x(A)AG(1_A) \quad (16.11)
\]
More complex “nonlinear” aggregation functions can be constructed by defining the aggregation function on all vertices of the hypercube. I.e., for each \(1_A : A \subseteq E\) we might have (for all \(A \subseteq E\)):

\[
AG(1_A) = w_A
\]  

(16.10)

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

\[
AG(x) = \sum_{A \subseteq E} x(A) w_A = \sum_{A \subseteq E} x(A) AG(1_A)
\]

(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

\[
WAVG_w(x) = AG_{w'}(x)
\]

(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 $1_A : A \subseteq E$ we might have (for all $A \subseteq E$):

$$AG(1_A) = w_A \quad (16.10)$$

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

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

- 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

$$WAVG_{w'}(x) = AG_w(x) \quad (16.12)$$

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

Prof. Jeff Bilmes EE563/Spring 2018/Submodularity - Lecture 16 - May 21st, 2018 F16/54 (pg.30/179)
Set function $f : 2^E \rightarrow \mathbb{R}$ is called a capacity if it is monotone non-decreasing, i.e., $f(A) \leq f(B)$ whenever $A \subseteq B$. 

Prof. Jeff Bilmes
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 $f : \{0, 1\}^m \to \mathbb{R}$.
Set function \( f : 2^E \rightarrow \mathbb{R} \) 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 \( f : \{0, 1\}^m \rightarrow \{0, 1\} \) and is a **pseudo-Boolean function** if \( f : \{0, 1\}^m \rightarrow \mathbb{R} \).

Any set function corresponds to a pseudo-Boolean function. I.e., given \( f : 2^E \rightarrow \mathbb{R} \), form \( f_b : \{0, 1\}^m \rightarrow \mathbb{R} \) as \( f_b(x) = f(A_x) \) where the \( A, x \) bijection is \( A = \{ e \in E : x_e = 1 \} \) and \( x = 1_A \).
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 $f_b(x) = f(Ax)$ where the $A, x$ bijection is $A = \{e \in E : x_e = 1\}$ and $x = 1_A$. Also, if we have an expression for $f_b$ we can construct a set function $f$ as $f(A) = f_b(1_A)$. We can also often relax $f_b$ to any $x \in [0, 1]^m$. 

Integration, Aggregation, and Weighted Averages
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 $x = 1_A$.

- Also, if we have an expression for $f_b$ we can construct a set function $f$ as $f(A) = f_b(1_A)$. 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 $f : 2^E \rightarrow \mathbb{R}$ 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 $f : \{0, 1\}^m \rightarrow \{0, 1\}$ and is a pseudo-Boolean function if $f : \{0, 1\}^m \rightarrow \mathbb{R}$.

- Any set function corresponds to a pseudo-Boolean function. I.e., given $f : 2^E \rightarrow \mathbb{R}$, form $f_b : \{0, 1\}^m \rightarrow \mathbb{R}$ as $f_b(x) = f(A_x)$ where the $A, x$ bijection is $A = \{e \in E : x_e = 1\}$ and $x = 1_A$.

- Also, if we have an expression for $f_b$ we can construct a set function $f$ as $f(A) = f_b(1_A)$. 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.
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

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

(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:

\[
C_f(w) = \sum_{i=1}^{m} w(e_i)(f(E_i) - f(E_{i-1}))
\]

(16.14)

where \( E_0 \triangleq \emptyset \).
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

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

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 $E_i = \{e_1, e_2, \ldots, e_i\}$.

- 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

$$\sum_{k=m}^{n} a_k b_k = \sum_{k=m}^{n} A_k (b_k - b_{k+1}) + A_n b_{n+1} - A_{m-1} b_m \quad (16.15)$$
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 $w_{e_1}$.

\[
\begin{align*}
  w(e_m) & \quad w(e_{m-1}) & \cdots & \quad w(e_5) & \quad w(e_4) & \quad w(e_3) & \quad w(e_2) & \quad w(e_1) \\
\end{align*}
\]
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 $w_{e_1}$.

$$w(e_m) \ w(e_{m-1}) \ \cdots \ \ w(e_5) \ w(e_4) \ w(e_3) \ w(e_2) \ w(e_1)$$

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

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

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

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

- 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 $w_{m+1} \triangleq 0$.

$$\tilde{f}(w) \triangleq \int_0^\infty F(\alpha) d\alpha \quad (16.17)$$
Now consider the integral, with \( w \in \mathbb{R}^E_+ \), and normalized \( f \) so that \( f(\emptyset) = 0 \). Recall \( w_{m+1} \stackrel{\text{def}}{=} 0 \).

\[
\tilde{f}(w) \stackrel{\text{def}}{=} \int_0^\infty F(\alpha) d\alpha
\]

\[
= \int_0^\infty f(\{e \in E : w_e > \alpha\}) d\alpha
\]
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 \( w_{m+1} \overset{\text{def}}{=} 0 \).

\[
\tilde{f}(w) \overset{\text{def}}{=} \int_0^\infty F(\alpha) d\alpha
\]

\[
= \int_0^\infty f(\{ e \in E : w_e > \alpha \}) d\alpha \quad (16.17)
\]

\[
= \int_{w_{m+1}}^\infty f(\{ e \in E : w_e > \alpha \}) d\alpha \quad (16.18)
\]
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 \( w_{m+1} \overset{\text{def}}{=} 0 \).

\[
\tilde{f}(w) \overset{\text{def}}{=} \int_0^\infty F(\alpha) d\alpha
\]

\[
= \int_0^\infty f(\{e \in E : w_e > \alpha\}) d\alpha
\]

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

\[
= \sum_{i=1}^{m} \int_{w_i}^{w_i+1} f(\{e \in E : w_e > \alpha\}) d\alpha
\]
Now consider the integral, with $w \in \mathbb{R}^E_+$, and normalized $f$ so that $f(\emptyset) = 0$. Recall $w_{m+1} \overset{\text{def}}{=} 0$.

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

(16.17)

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

(16.18)

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

(16.19)

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

(16.20)

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

(16.21)
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:

\[
\tilde{f}(w) \overset{\text{def}}{=} \int_0^\infty F(\alpha) d\alpha \quad (16.22)
\]

where the function \( F \) is defined as before.
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:

$$\tilde{f}(w) \overset{\text{def}}{=} \int_0^\infty F(\alpha) d\alpha$$  \hspace{1cm} (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:

$$\tilde{f}(w) \overset{\text{def}}{=} \int_0^\infty F(\alpha) d\alpha$$  \hspace{1cm} (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.
Recall, we want to produce some notion of generalized aggregation function having the flavor of:

\[
AG(x) = \sum_{A \subseteq E} x(A)w_A = \sum_{A \subseteq E} x(A)AG(1_A)
\]  

(16.23)

how does this correspond to Lovász extension?
Recall, we want to produce some notion of generalized aggregation function having the flavor of:

$$AG(x) = \sum_{A \subseteq E} x(A)w_A = \sum_{A \subseteq E} x(A)AG(1_A)$$  \hspace{1cm} (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 $V_1, V_2, \ldots, V_q$. $V_i$
Recall, we want to produce some notion of generalized aggregation function having the flavor of:

\[
AG(x) = \sum_{A \subseteq E} x(A)w_A = \sum_{A \subseteq E} x(A)AG(1_A)
\]

(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 \(V_1, V_2, \ldots, V_q\).

E.g., for each \(i\), \(V_i = \{1_{A_1}, 1_{A_2}, \ldots, 1_{A_k}\}\) (\(k\) vertices) and the convex hull of \(V_i\) defines the \(i^{th}\) polytope.
Recall, we want to produce some notion of generalized aggregation function having the flavor of:

$$AG(x) = \sum_{A \subseteq E} x(A)w_A = \sum_{A \subseteq E} x(A)AG(1_A)$$ (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 $V_1, V_2, \ldots, V_q$.

E.g., for each $i$, $V_i = \{1_{A_1}, 1_{A_2}, \ldots, 1_{A_k}\}$ ($k$ vertices) and the convex hull of $V_i$ defines the $i^{th}$ polytope.

This forms a “triangulation” of the hypercube.
Recall, we want to produce some notion of generalized aggregation function having the flavor of:

\[
AG(x) = \sum_{A \subseteq E} x(A)w_A = \sum_{A \subseteq E} x(A)AG(1_A)
\]  

(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 $i$, $\mathcal{V}_i = \{1_{A_1}, 1_{A_2}, \ldots, 1_{A_k}\}$ ($k$ vertices) and the convex hull of $\mathcal{V}_i$ 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 $x \in \text{conv} (\mathcal{V}(x))$. 
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 $1_A \in \text{conv}(\mathcal{V}(x))$. 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.
Most generally, for $x \in [0, 1]^m$, let us define the (unique) coefficients $\alpha^x_0(A)$ and $\alpha^x_i(A)$ that define the affine transformation of the coefficients of $x$ to be used with the particular hypercube vertex $1_A \in \text{conv}(\mathcal{V}(x))$. The affine transformation is as follows:

$$\alpha^x_0(A) + \sum_{j=1}^{m} \alpha^x_j(A)x_j \in \mathbb{R}$$ \hspace{1cm} (16.24)

Note that many of these coefficients are often zero.

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

$$\text{AG}(x) \overset{\text{def}}{=} \sum_{A: 1_A \in \mathcal{V}(x)} \left( \alpha^x_0(A) + \sum_{j=1}^{m} \alpha^x_j(A)x_j \right) \text{AG}(1_A)$$ \hspace{1cm} (16.25)
We can define a canonical triangulation of the hypercube in terms of permutations of the coordinates. I.e., given some permutation $\sigma$, define

$$
\text{conv}(V_{\sigma}) = \{x \in [0, 1]^n | x_{\sigma(1)} \geq x_{\sigma(2)} \geq \cdots \geq x_{\sigma(m)}\} \quad (16.26)
$$

Then these $m!$ blocks of the partition are called the canonical partitions of the hypercube.
We can define a canonical triangulation of the hypercube in terms of permutations of the coordinates. I.e., given some permutation $\sigma$, define

$$\text{conv}(V_\sigma) = \{x \in [0, 1]^n | x_{\sigma(1)} \geq x_{\sigma(2)} \geq \cdots \geq x_{\sigma(m)}\}$$  \hspace{1cm} (16.26)$$

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

With this, we can define $\{V_i\}_{i=1}^{m!}$ as the vertices of $\text{conv}(V_\sigma)$ for each permutation $\sigma$. 
We can define a canonical triangulation of the hypercube in terms of permutations of the coordinates. I.e., given some permutation \( \sigma \), define

\[
\text{conv}(V_\sigma) = \{ x \in [0, 1]^n | x_{\sigma(1)} \geq x_{\sigma(2)} \geq \cdots \geq x_{\sigma(m)} \} \tag{16.26}
\]

Then these \( m! \) blocks of the partition are called the canonical partitions of the hypercube.

With this, we can define \( \{V_i\}_i \) as the vertices of \( \text{conv}(V_\sigma) \) for each permutation \( \sigma \). In this case, we have:
We can define a canonical triangulation of the hypercube in terms of permutations of the coordinates. I.e., given some permutation $\sigma$, define

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

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

With this, we can define $\{V_i\}_i$ as the vertices of $\text{conv}(V_\sigma)$ for each permutation $\sigma$. 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 $\alpha^x_0(A) + \sum_{j=1}^{m} \alpha^x_j(A) x_j = x_{\sigma_i} - x_{\sigma_{i-1}}$ for $A = E_i = \{e_{\sigma_1}, \ldots, e_{\sigma_i}\}$ for appropriate order $\sigma$. 

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

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

$$\alpha^x_0(A) + \sum_{j=1}^{m} \alpha^x_j(A)x_j = x_{\sigma_i} - x_{\sigma_{i-1}}$$

for $A = E_i = \{e_{\sigma_1}, \ldots, e_{\sigma_i}\}$ for appropriate order $\sigma$.

Hence, Lovász extension is a generalized aggregation function.
\[ f(a) = \frac{1}{n} \sum_{i=1}^{n} x_i \]

\[ f(x) = \frac{1}{n} \sum_{i=1}^{n} \sum x_i \]

\[ f(a) = \text{sample} (A) \]

\[ f(x) = \text{sample} (x) \]
Lovász extension as max over orders

We can also write the Lovász extension as follows:

\[
\tilde{f}(w) = \max_{\sigma \in \Pi[m]} w^\top c^\sigma
\]  

(16.27)

where \( \Pi[m] \) is the set of \( m! \) permutations of \([m] = E\), \( \sigma \in \Pi[m] \) is a particular permutation, and \( c^\sigma \) is a vector associated with permutation \( \sigma \) defined as:

\[
c^\sigma_i = f(E_{\sigma_i}) - f(E_{\sigma_{i-1}})
\]  

(16.28)

where \( E_{\sigma_i} = \{e_{\sigma_1}, e_{\sigma_2}, \ldots, e_{\sigma_i}\} \).
We can also write the Lovász extension as follows:

\[ \tilde{f}(w) = \max_{\sigma \in \Pi[m]} w^\top c^\sigma \]  

(16.27)

where \( \Pi[m] \) is the set of \( m! \) permutations of \( [m] = E \), \( \sigma \in \Pi[m] \) is a particular permutation, and \( c^\sigma \) is a vector associated with permutation \( \sigma \) defined as:

\[ c^\sigma_i = f(E_{\sigma_i}) - f(E_{\sigma_{i-1}}) \]  

(16.28)

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

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

\[ \tilde{f}(w) = \max_{x \in P_f} w^\top x = \max_{x \in B_f} w^\top x \]  

(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, let's use \( \{ w \geq \alpha \} \equiv \{ e \in E : w(e) \geq \alpha \} \), called the weak \( \alpha \)-sup-level set of \( w \).
Lovász extension, defined in multiple ways

As shorthand notation, let's use \( \{ w \geq \alpha \} \equiv \{ e \in E : w(e) \geq \alpha \} \), called the weak \( \alpha \)-sup-level set of \( w \). A similar definition holds for \( \{ w > \alpha \} \) (called the strong \( \alpha \)-sup-level set of \( w \)).
Lovász extension, defined in multiple ways

- As shorthand notation, let's use \( \{ w \geq \alpha \} \equiv \{ e \in E : w(e) \geq \alpha \} \), called the weak \( \alpha \)-sup-level set of \( w \). A similar definition holds for \( \{ w > \alpha \} \) (called the strong \( \alpha \)-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, let's use \( \{ w \geq \alpha \} \equiv \{ e \in E : w(e) \geq \alpha \} \), called the weak \( \alpha \)-sup-level set of \( w \). A similar definition holds for \( \{ w > \alpha \} \) (called the strong \( \alpha \)-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, let's use \( \{w \geq \alpha\} \equiv \{e \in E : w(e) \geq \alpha\} \), called the weak \( \alpha \)-sup-level set of \( w \). A similar definition holds for \( \{w > \alpha\} \) (called the strong \( \alpha \)-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:

\[
\tilde{f}(w) = \sum_{i=1}^m w(e_i) f(e_i | E_{i-1})
\]

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

\[
= \sum_{i=1}^{m-1} \lambda_i f(E_i)
\]

(16.30)
Lovász extension, as integral

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

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

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

\[
= \int_{\min \{w_1, \ldots, w_m\}}^{+\infty} f\{w \geq \alpha\}d\alpha + f(E)\min \{w_1, \ldots, w_m\} \quad (16.35)
\]

\[
= \int_{0}^{+\infty} f\{w \geq \alpha\}d\alpha + \int_{-\infty}^{0} [f\{w \geq \alpha\} - f(E)]d\alpha \quad (16.36)
\]
In fact, we have that, given function $f$, and any $w \in \mathbb{R}^E$:

$$\tilde{f}(w) = \int_{-\infty}^{+\infty} \hat{f}(\alpha) d\alpha$$  \hspace{1cm} (16.37)

where

$$\hat{f}(\alpha) = \begin{cases} f(\{w \geq \alpha\}) & \text{if } \alpha \geq 0 \\ f(\{w \geq \alpha\}) - f(E) & \text{if } \alpha < 0 \end{cases}$$  \hspace{1cm} (16.38)
In fact, we have that, given function $f$, and any $w \in \mathbb{R}^E$:

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

where

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

So we can write it as a simple integral over the right function.
In fact, we have that, given function \( f \), and any \( w \in \mathbb{R}^E \):

\[
\tilde{f}(w) = \int_{-\infty}^{+\infty} \hat{f}(\alpha) d\alpha
\]

(16.37)

where

\[
\hat{f}(\alpha) = \begin{cases} 
    f(\{w \geq \alpha\}) & \text{if } \alpha \geq 0 \\
    f(\{w \geq \alpha\}) - f(E) & \text{if } \alpha < 0
\end{cases}
\]

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

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

we may use open intervals since sets of zero measure don’t change integration.
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 $\alpha$, we have

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

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

Inside the integral, then, this recovers Eqn. (16.34).
To show Eqn. (16.36), start with Eqn. (16.35), note

\[ w_m = \min \{ w_1, \ldots, w_m \}, \text{ take any } \beta \leq \min \{ 0, w_1, \ldots, w_m \}, \text{ and form:} \]

\[ \tilde{f}(w) \]
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:

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

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

\[ w_m = \min \{ w_1, \ldots, w_m \}, \text{ take any } \beta \leq \min \{ 0, w_1, \ldots, w_m \}, \text{ and form:} \]

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

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

\[ \]
To show Eqn. (16.36), start with Eqn. (16.35), note

\[ w_m = \min \{ w_1, \ldots, w_m \}, \text{ take any } \beta \leq \min \{ 0, w_1, \ldots, w_m \}, \text{ and form:} \]

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

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

\[ = \int_{\beta}^{+\infty} f(\{ w \geq \alpha \}) d\alpha - \int_{\beta}^{w_m} f(E) d\alpha + \int_{0}^{w_m} f(E) d\alpha \]
To show Eqn. (16.36), start with Eqn. (16.35), note

\[ w_m = \min \{w_1, \ldots, w_m\}, \]  

and take any \( \beta \leq \min \{0, w_1, \ldots, w_m\} \), and form:

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

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

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

\[ = \int_{0}^{+\infty} f(\{w \geq \alpha\})d\alpha + \int_{0}^{\beta} f(\{w \geq \alpha\})d\alpha - \int_{\beta}^{0} f(E)d\alpha \]
To show Eqn. (16.36), start with Eqn. (16.35), note

\[ w_m = \min \{ w_1, \ldots, w_m \}, \text{ take any } \beta \leq \min \{ 0, w_1, \ldots, w_m \}, \text{ and form:} \]

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

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

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

\[
= \int_{0}^{+\infty} f(\{ w \geq \alpha \}) \, d\alpha + \int_{\beta}^{0} [f(\{ w \geq \alpha \}) - f(E)] \, d\alpha
\]
Lovász extension, as integral

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

\[ w_m = \min \{ w_1, \ldots, w_m \}, \text{ take any } \beta \leq \min \{ 0, w_1, \ldots, w_m \}, \text{ and form:} \]

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

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

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

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

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

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

\[ w_m = \min \{ w_1, \ldots, w_m \}, \text{ take any } \beta \leq \min \{ 0, w_1, \ldots, w_m \}, \text{ and form:} \]

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

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

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

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

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 $\tilde{f}$ is the Lovász extension of $f$ for $x \in \mathbb{R}$.

2. If $w \in \mathbb{R}^E + 1$ then $\tilde{f}(w) = \int_{\{w\}} f(\{w\}) \, d\mu$.

3. For $w \in \mathbb{R}^E$, and $\mu \in \mathbb{R}$, we have $\tilde{f}(w + \mu) = \tilde{f}(w) + \mu f(E)$.

4. Positive homogeneity: I.e., $\tilde{f}(\mu w) = \mu \tilde{f}(w)$ for $\mu \geq 0$.

5. For all $A \subseteq E$, $\tilde{f}(1_A) = f(A)$.

6. $f$ symmetric as in $f(A) = f(E \cap A)$, 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 $w = \sum_{i=1}^k w_i$ with $w_i \in E_i$ for $1 \leq i \leq k$, and with $E_1: i = E_1 \cup E_2 \cup \cdots \cup E_i$, then $\tilde{f}(w) = \sum_{i=1}^k f(E_i | E_1:i) = \sum_{i=1}^k f(E_1:i)$.
Lovász extension properties

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

Theorem 16.5.1

Let $f, g : 2^E \rightarrow \mathbb{R}$ 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 \( \lambda f \) 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 \( \lambda f \) for \( \lambda \in \mathbb{R} \).

2. **If \( w \in \mathbb{R}_+^E \) then** \( \tilde{f}(w) = \int_0^{+\infty} f\left(\{w \geq \alpha\}\right) d\alpha \).
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 \( \lambda f \) 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 \( \tilde{f}(w + \alpha 1_E) = \tilde{f}(w) + \alpha f(E) \).
Lovász extension properties

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

**Theorem 16.5.1**

Let \( f, g : 2^E \rightarrow \mathbb{R} \) 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 \( \lambda f \) 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 \( \tilde{f}(w + \alpha 1_E) = \tilde{f}(w) + \alpha f(E) \).

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 \( f, g : 2^E \rightarrow \mathbb{R} \) 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 \( \lambda f \) 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 \( \tilde{f}(w + \alpha 1_E) = \tilde{f}(w) + \alpha f(E) \).

4. **Positive homogeneity:** i.e., \( \tilde{f}(\alpha w) = \alpha \tilde{f}(w) \) for \( \alpha \geq 0 \).

5. For all \( A \subseteq E \), \( \tilde{f}(1_A) = f(A) \).
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 \( \lambda f \) 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** \( \tilde{f}(w + \alpha 1_E) = \tilde{f}(w) + \alpha f(E) \).

4. **Positive homogeneity**: i.e., \( \tilde{f}(\alpha w) = \alpha \tilde{f}(w) \) for \( \alpha \geq 0 \).

5. **For all** \( A \subseteq E \), \( \tilde{f}(1_A) = f(A) \).

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

4. **Positive homogeneity**: I.e., $\tilde{f}(\alpha w) = \alpha \tilde{f}(w)$ for $\alpha \geq 0$.

5. For all $A \subseteq E$, $\tilde{f}(1_A) = f(A)$.

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 $w = \sum_{i=1}^{k} \gamma_i 1_{E_k}$ 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

$$\tilde{f}(w) = \sum_{i=1}^{k} \gamma_i f(E^i | E^{1:i-1}) = \sum_{i=1}^{k-1} f(E^{1:i})(\gamma_i - \gamma_{i+1}) + f(E)^{\gamma_k}.$$
Consider property property 3, which says that
\[ \tilde{f}(w + \alpha 1_E) = \tilde{f}(w) + \alpha f(E). \]
Consider property property 3, for example, which says that
\[ \tilde{f}(w + \alpha 1_E) = \tilde{f}(w) + \alpha f(E). \]

This means that, say when \( m = 2 \), that as we move along the line \( w_1 = w_2 \), the Lovász extension scales linearly.
Consider property property 3, for example, which says that
\[ \tilde{f}(w + \alpha 1_E) = \tilde{f}(w) + \alpha f(E). \]

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 \( 1_E \).
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

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

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 $f(\{w \leq \alpha\}) = f(\{w > \alpha\})$. 
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

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

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

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

\[
(\text{a}) \quad \int_{-\infty}^{\infty} f\{w \leq \alpha\} d\alpha
\]

\[
(16.42)
\]

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

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

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

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 $f(\{w \leq \alpha\}) = f(\{w > \alpha\})$. 
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

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

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

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

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 $f(\{w \leq \alpha\}) = f(\{w > \alpha\})$. 
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

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

\[
(\text{(a)}) = \int_{-\infty}^{\infty} f(\{w \leq \alpha\}) d\alpha \quad (\text{(b)}) = \int_{-\infty}^{\infty} f(\{w > \alpha\}) d\alpha
\]

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

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 \( f(\{w \leq \alpha\}) = f(\{w > \alpha\}) \).
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 $\alpha > w_1 \geq w_1$, 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 $\alpha$ as a uniform random variable on $[0, 1]$ and let $h(\alpha)$ be a function of $\alpha$. 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 $\tilde{f}(w) = \int_0^{w_1} f(\{w \geq \alpha\})d\alpha$
- 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 $\alpha$ as a uniform random variable on $[0, 1]$ and let $h(\alpha)$ be a function of $\alpha$. 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

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

where $\alpha$ 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 \( \alpha > w_1 \geq w \), 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 \( \alpha \) as a uniform random variable on \([0,1]\) and let \( h(\alpha) \) be a function of \( \alpha \). 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

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

where \( \alpha \) 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

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

  (16.44)  
  (16.45)
Simple expressions for Lovász E. with $m = 2, E = \{1, 2\}$

- If $w_1 \geq w_2$, then

\[
\tilde{f}(w) = w_1 f(\{1\}) + w_2 f(\{2\}|\{1\})
\]

\[
= (w_1 - w_2) f(\{1\}) + w_2 f(\{1, 2\})
\] (16.44)

\[
\tilde{f}(w) = w_1 f(\{1\}) + w_2 f(\{2\}|\{1\})
\]

\[
= (w_1 - w_2) f(\{1\}) + w_2 f(\{1, 2\})
\] (16.45)

- If $w_1 \leq w_2$, then

\[
\tilde{f}(w) = w_2 f(\{2\}) + w_1 f(\{1\}|\{2\})
\]

\[
= (w_2 - w_1) f(\{2\}) + w_1 f(\{1, 2\})
\] (16.46)

\[
\tilde{f}(w) = w_2 f(\{2\}) + w_1 f(\{1\}|\{2\})
\]

\[
= (w_2 - w_1) f(\{2\}) + w_1 f(\{1, 2\})
\] (16.47)
If $w_1 \geq w_2$, then

\[
\tilde{f}(w) = w_1 f(\{1\}) + w_2 f(\{2\}|\{1\}) \\
= (w_1 - w_2)f(\{1\}) + w_2 f(\{1, 2\}) \\
= \frac{1}{2} f(1)(w_1 - w_2) + \frac{1}{2} f(1)(w_1 - w_2) \\
+ \frac{1}{2} f(\{1, 2\})(w_1 + w_2) - \frac{1}{2} f(\{1, 2\})(w_1 - w_2) \\
+ \frac{1}{2} f(2)(w_1 - w_2) + \frac{1}{2} f(2)(w_2 - w_1)
\]
If $w_1 \geq w_2$, then

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

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

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

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

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

A similar (symmetric) expression holds when $w_1 \leq w_2$. 
This gives, for general $w_1, w_2$, that

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

(16.53)

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

(16.54)

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

(16.55)

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

(16.56)

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

(16.57)
This gives, for general $w_1, w_2$, that

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

$$+ \frac{1}{2} (f(\{1\}) - f(\{2\}) + f(\{1, 2\})) w_1$$  \hfill (16.54)

$$+ \frac{1}{2} (-f(\{1\}) + f(\{2\}) + f(\{1, 2\})) w_2$$  \hfill (16.55)

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

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

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

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

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

$$= -(f(\{1\}) + f(\{2\}) - f(\{1, 2\})) \min \{w_1, w_2\}$$  \hspace{1cm} (16.56)

$$+ f(\{1\})w_1 + f(\{2\})w_2$$  \hspace{1cm} (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

\[
\tilde{f}(w) = w_1 f(\{1\}) + w_2 f(\{2\}|\{1\})
\]  

(16.58)
Example: $m = 2$, $E = \{1, 2\}$, contours

- If $w_1 \geq w_2$, then

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

- If $w = (1, 0) / f(\{1\}) = \left(1/f(\{1\}), 0\right)$ then $\tilde{f}(w) = 1$. 
Example: \( m = 2, \ E = \{1, 2\} \), contours

- If \( w_1 \geq w_2 \), then

\[
\tilde{f}(w) = w_1 f(\{1\}) + w_2 f(\{2\}|\{1\})
\]  

(16.58)

- If \( w = (1, 0)/f(\{1\}) = \left(\frac{1}{f(\{1\})}, 0\right) \) 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
  \[ \tilde{f}(w) = w_1 f(\{1\}) + w_2 f(\{2\}|\{1\}) \]  \hspace{1cm} (16.58)

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

- If $w_1 \leq w_2$, then
  \[ \tilde{f}(w) = w_2 f(\{2\}) + w_1 f(\{1\}|\{2\}) \]  \hspace{1cm} (16.59)
Example: $m = 2$, $E = \{1, 2\}$, contours

- If $w_1 \geq w_2$, then

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

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

- If $w_1 \leq w_2$, then

  $$\tilde{f}(w) = w_2 f(\{2\}) + w_1 f(\{1\} | \{2\})$$  \hspace{1cm} (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
  \[ \tilde{f}(w) = w_1 f(\{1\}) + w_2 f(\{2\}|\{1\}) \]  
  (16.58)

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

- If $w_1 \leq w_2$, then
  \[ \tilde{f}(w) = w_2 f(\{2\}) + w_1 f(\{1\}|\{2\}) \]  
  (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 $\tilde{f}(w) = 1$. 

Prof. Jeff Bilmes
Example: $m = 2$, $E = \{1, 2\}$, contours

- If $w_1 \geq w_2$, then
  \[
  \tilde{f}(w) = w_1 f(\{1\}) + w_2 f(\{2\} \mid \{1\})
  \]  
  (16.58)

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

- If $w_1 \leq w_2$, then
  \[
  \tilde{f}(w) = w_2 f(\{2\}) + w_1 f(\{1\} \mid \{2\})
  \]  
  (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 $\{w \in \mathbb{R}^2 : \tilde{f}(w) = 1\}$, particular marked points of form $w = 1_A \times \frac{1}{f(A)}$ for certain $A$, where $\tilde{f}(w) = 1$. 
Example: $m = 2, \ E = \{1, 2\}$

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

\[
\begin{align*}
(0, 1)/f(\{2\}) & \quad (1, 1)/f(\{1, 2\}) \\
(1, 0)/f(\{1\}) & \quad w_1 > w_2
\end{align*}
\]

\[w_2 > w_1\]
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 $\tilde{f}(w + \alpha 1_E) = \tilde{f}(w) + \alpha f(E)$, we have that $\tilde{f}(w + \alpha 1_E) = \tilde{f}(w)$ 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 \( \tilde{f}(w + \alpha 1_E) = \tilde{f}(w) + \alpha f(E) \), we have that \( \tilde{f}(w + \alpha 1_E) = \tilde{f}(w) \) 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 \( \left\{ x : x = \alpha 1_E, \, \alpha > 0 \right\} \) since moving in direction \( 1_E \) changes nothing.
Example: \( m = 3, \ E = \{1, 2, 3\} \)

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

- Example 1 (from Bach-2011):  
  \[
  f(A) = 1_{|A| \in \{1, 2\}} = \min \{|A|, 1\} + \min \{|E \setminus A|, 1\} - 1
  \]
  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): $f(A) = |1_{1 \in A} - 1_{2 \in A}| + |1_{2 \in A} - 1_{3 \in A}|$
Example: $m = 3$, $E = \{1, 2, 3\}$

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

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 $M_j = \sum_{i=1}^{j} m(e_i)$
Lovász extension

Example: Lovász extension of concave over modular

- Let $m : E \rightarrow \mathbb{R}_+$ 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

\[
\tilde{f}(w) = \sum_{i=1}^m w(e_i)(g(M_i) - g(M_{i-1}))
\] (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

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

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

\[
\tilde{f}(w) = \sum_{i=1}^{m} w(e_i) (g(i) - g(i - 1)) \tag{16.61}
\]
Example: Lovász extension and cut functions

- Cut Function: Given a non-negative weighted graph $G = (V, E, m)$ where $m : E \rightarrow \mathbb{R}_+$ is a modular function over the edges, we know from Lecture 2 that $f : 2^V \rightarrow \mathbb{R}_+$ 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 $m : E \rightarrow \mathbb{R}_+$ is a modular function over the edges, we know from Lecture 2 that $f : 2^V \rightarrow \mathbb{R}_+$ 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 $m_{ij} = m((i, j))$:

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

---

**Prof. Jeff Bilmes**
EE563/Spring 2018/Submodularity - Lecture 16 - May 21st, 2018

F43/54 (pg.143/179)
Example: Lovász extension and cut functions

- Cut Function: Given a non-negative weighted graph $G = (V, E, m)$ where $m : E \rightarrow \mathbb{R}_+$ is a modular function over the edges, we know from Lecture 2 that $f : 2^V \rightarrow \mathbb{R}_+$ 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, $w_1 \geq w_2 \geq \cdots \geq w_n$. 
Example: Lovász extension and cut functions

- **Cut Function:** Given a non-negative weighted graph \( G = (V, E, m) \) where \( m : E \rightarrow \mathbb{R}_+ \) is a modular function over the edges, we know from Lecture 2 that \( f : 2^V \rightarrow \mathbb{R}_+ \) 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, \( w_1 \geq w_2 \geq \cdots \geq w_n \).

- **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 \( W_k \triangleq \sum_{i=1}^{k} w(e_i) \).

<table>
<thead>
<tr>
<th>( f(A) )</th>
<th>( \tilde{f}(w) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(</td>
<td>A</td>
</tr>
<tr>
<td>( \min(</td>
<td>A</td>
</tr>
<tr>
<td>( \min(</td>
<td>A</td>
</tr>
<tr>
<td>( \min(</td>
<td>A</td>
</tr>
<tr>
<td>( \min(</td>
<td>A</td>
</tr>
<tr>
<td>( \min(</td>
<td>A</td>
</tr>
</tbody>
</table>

(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:

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

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:

$$\min_{w^1, \ldots, w^k \in \mathbb{R}^n} \sum_{j=1}^k \frac{1}{m} \sum_{i=1}^m \ell(y^k_i, (w^k)^T x_i) + \lambda \Omega(w^k), \quad (16.65)$$

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

$$\min_{x_1, \ldots, x_m} \min_{w^1, \ldots, w^k \in \mathbb{R}^n} \sum_{j=1}^k \frac{1}{m} \sum_{i=1}^m \ell(y^k_i, (w^k)^T x_i) + \lambda \Omega(w^k), \quad (16.66)$$
Norms, sparse norms, and computer vision

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

- $1$-norm promotes sparsity (prefer solutions with zero entries).

- Image denoising, total variation is useful, norm takes form:
  
  \[
  \|w\|_1 = \sum_{i=2}^{N} |w_i - w_{i-1}|
  \] (16.67)

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

  (Rodriguez, 2009)
Common norms include $p$-norm $\Omega(w) = \|w\|_p = (\sum_{i=1}^{p} w_i^p)^{1/p}$

1-norm promotes sparsity (prefer solutions with zero entries).
Norms, sparse norms, and computer vision

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

$$\Omega(w) = \sum_{i=2}^{N} |w_i - w_{i-1}|$$  \hspace{1cm} (16.67)
Norms, sparse norms, and computer vision

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

$$\Omega(w) = \sum_{i=2}^{N} |w_i - w_{i-1}|$$ (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 \( w \in \mathbb{R}^V \), supp\((w) \in \{0,1\}^V \) has 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, $\|w\|_0 = 1^T \text{supp}(w)$. 

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 $w \in \mathbb{R}^V$, supp$(w) \in \{0, 1\}^V$ has supp$(w)(v) = 1$ iff $w(v) > 0$
- Desirable sparse norm: count the non-zero values, $\|w\|_0 = 1^T$ 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, \( \|w\|_0 = 1^T \text{supp}(w) \).
- 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 $w \in \mathbb{R}^V$, supp$(w) \in \{0, 1\}^V$ has supp$(w)(v) = 1$ iff $w(v) > 0$
- Desirable sparse norm: count the non-zeros, $\|w\|_0 = 1^T \text{supp}(w)$.
- 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 $f : 2^V \rightarrow \mathbb{R}_+$, $f(\text{supp}(w))$ 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, $\|w\|_0 = 1^T \text{supp}(w)$.
- 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 $f : 2^V \rightarrow \mathbb{R}_+$, $f(\text{supp}(w))$ 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, $\|w\|_0 = 1^T \text{supp}(w)$.
- 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 $f : 2^V \rightarrow \mathbb{R}_+$, $f(\text{supp}(w))$ 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!
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, $\|w\|_0 = 1^T \text{supp}(w)$.
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 $f : 2^V \rightarrow \mathbb{R}^+$, $f(\text{supp}(w))$ 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 $\|w\|_f = \tilde{f}(|w|)$, renders the function symmetric about all orthants (i.e., $\|w\|_f = \|b \odot w\|_f$ 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 $\ell_1$ norm, and the Lovász extension of the modular function $f(A) = m(A)$ is the weighted $\ell_1$ 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 $\ell_2$ 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 $\ell_1$ norm, and the Lovász extension of the modular function $f(A) = m(A)$ is the weighted $\ell_1$ 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 $\ell_2$ 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 $\ell_1$ norm, and the Lovász extension of the modular function $f(A) = m(A)$ is the weighted $\ell_1$ 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 $\ell_2$ 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:

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

(16.68)

where \( \triangle^n(x) = \{ p \in \mathbb{R}^{2^n} : \sum_{S \subseteq V} p_S = 1, p_S \geq 0 \forall S \subseteq V, \ & \sum_{S \subseteq V} p_S 1_S = x \} \)
The concave closure is defined as:

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

(16.68)

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

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

\[
\hat{f}(x) = \max_{p \in \Delta^n(x)} \sum_{S \subseteq V} p_S f(S)
\]

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

- This is tight at the hypercube vertices, concave, and the concave envelope 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.
The concave closure is defined as:

\[
\hat{f}(x) = \max_{p \in \Delta^n(x)} \sum_{S \subseteq V} p_S f(S)
\]  

(16.68)

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

This is tight at the hypercube vertices, concave, and the concave envelope 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 maximization 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 $x \in [0, 1]^V = [0, 1]^n$

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

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

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

- 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 $\Pr(u \in S') = x_u, v \neq u$. 
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 \( f(1_A) \) 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 \( f(1_A) \) 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., \( \tilde{f}(x_1, x_2, \ldots, ax_k + \beta x_k', \ldots, x_n) = \alpha \tilde{f}(x_1, x_2, \ldots, x_k, \ldots, x_n) + \beta \tilde{f}(x_1, x_2, \ldots, x_k', \ldots, x_n) \))
Rather than the concave closure, multi-linear extension is used as a surrogate. For $x \in [0, 1]^V = [0, 1]^n$

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

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 $f(1_A)$ 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.
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.,
  \[ \frac{\partial^2 \tilde{f}}{\partial x_i \partial x_j} \leq 0 \text{ for all } i, j \in V \text{ within } [0, 1]^V. \]

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).

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

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

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

$$= E_{S \sim x}[f(S + v)] - E_{S \sim x}[f(S - v)]$$  \hspace{1cm} (16.72)

$$\geq 0$$  \hspace{1cm} (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 $i \in V.$
Second part of proof (antitone supergradient) also relies on simple consequence of multilinearity, namely multilinearity of the derivative as well. In this case

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

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

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

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

\[
\leq 0 \tag{16.78}
\]

since by submodularity, we have

\[
f(S + i - j) + f(S - i + j) \geq f(S + i + j) + f(S - i - j) \tag{16.79}
\]
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 $\alpha$ for any $z \in \mathbb{R}_+$).

- If $f$ is submodular then $\tilde{f}$ 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 $\alpha$ for any $u \neq v$).