Submodular Functions, Optimization, and Applications to Machine Learning
— Fall Quarter, Lecture 17 —

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

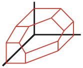
Prof. Jeff Bilmes

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

Nov 30th, 2020

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

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



Class Road Map - EE563

- L1(9/30): Motivation, Applications, Definitions, Properties
- L2(10/5): Sums concave(modular), uses (diversity/costs, feature selection), information theory
- L3(10/7): Monge, More Definitions, Graph and Combinatorial Examples,
- L4(10/12): Graph & Combinatorial Examples, Matrix Rank, Properties, Other Defs, Independence
- L5(10/14): Properties, Defs of Submodularity, Independence
- L6(10/19): Matroids, Matroid Examples, Matroid Rank,
- L7(10/21): Matroid Rank, More on Partition Matroid, Laminar Matroids, System of Distinct Reps, Transversals
- L8(10/26): Transversal Matroid, Matroid and representation, Dual Matroid
- L9(10/28): Other Matroid Properties, Combinatorial Geometries, Matroid and Greedy, Polyhedra, Matroid Polytopes
- L10(11/2): Matroid Polytopes, Matroids $\rightarrow$ Polymatroids
- L11(11/4): Matroids $\rightarrow$ Polymatroids, Polymatroids
- L12(11/9): Polymatroids, Polymatroids and Greedy
- L–(11/11): Veterans Day, Holiday
- L13(11/16): Polymatroids and Greedy, Possible Polytopes, Extreme Points, Cardinality Constrained Maximization
- L14(11/18): Cardinality Constrained Maximization, Curvature
- L15(11/23): Curvature, Submodular Max w. Other Constraints, Start Cont. Extensions
- L16(11/25): Submodular Max w. Other Constraints, Cont. Extensions, Lovász extension
- L17(11/30): Choquet Integration, Non-linear Measure/Aggregation, Definitions/Properties, Examples.
- L18(12/2):
- L19(12/7):
- L20(12/9):
- L21(12/14): final meeting (presentations) maximization.

Last day of instruction, Fri. Dec 11th. Finals Week: Dec 12-18, 2020

Rest of class

- Homework 4 posted, due Thursday Dec 17th, 2020, 11:55pm.
- Final project paper proposal, due Sunday Dec 6th, 11:59pm.
- Final project 4-page paper and presentation slides, due Sunday Dec 13th, 11:59pm.
- Final project presentation, Monday Dec 14th, starting at 10:30am.
- Final project: Read and present a recent (past 5 years) paper on submodular/supermodular optimization. Paper should have both a theoretical and practical component. What is due: (1) 4-page paper summary, and (2) 10 minute presentation about the paper, will be giving presentations on Monday 12/14/2020. You must choose your paper before the 14th (this will be HW5), and you must turn in your slides and 4-page paper (this will be HW6).
- Recall, grades will be based on a combination of a final project (40%) and the four homeworks (60%).

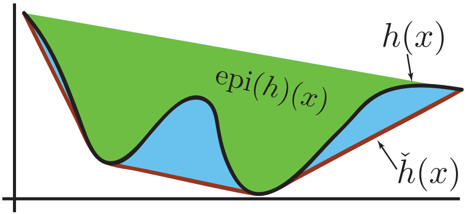
Def: Convex Envelope of a function

- Given any function $h : \mathcal{D}_h \to \mathbb{R}$, where $\mathcal{D}_h \subseteq \mathbb{R}^n$, define the new function $\check{h} : \mathbf{R}^n \to \mathbb{R}$ via:

$$\check{h}(x) = \sup \{g(x) : g \text{ is convex } \& \ g(y) \le h(y), \forall y \in \mathcal{D}_h\} \tag{17.1}$$

- I.e., (1) $\check{h}(x)$ is convex, (2) $\check{h}(x) \le h(x), \forall x$, and (3) if $g(x)$ is any convex function having the property that $g(x) \le h(x), \forall x$, then $g(x) \le \check{h}(x)$.
- Alternatively,

$$\check{h}(x) = \inf \{t : (x,t) \in \text{convexhull}(\text{epigraph}(h))\} \tag{17.2}$$



Convex Closure of Discrete Set Functions

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

$$\check{f}(x) = \min_{p \in \triangle^n(x)} \sum_{S \subseteq V} p_S f(S) \tag{17.1}$$

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

- Hence, $\triangle^n(x)$ is the set of all probability distributions over the 2^n vertices of the hypercube, and where the expected value of the characteristic vectors of those points is equal to $x \in [0,1]^V$, i.e., for any $p \in \triangle^n(x)$, $E_{S \sim p}(\mathbf{1}_S) = \sum_{S \subseteq V} p_S \mathbf{1}_S = x$.
- Hence, $\check{f}(x) = \min_{p \in \triangle^n(x)} E_{S \sim p}[f(S)]$
- We will show that this is **a** convex extension. Does it have any special properties?

Convex Closure of Discrete Set Functions

- Given, $\check{f}(x) = \min_{p \in \triangle^n(x)} E_{S \sim p}[f(S)]$, we ~~can~~ did show:
  1. that $\check{f}$ is tight (i.e., $\forall S \subseteq V$, we have $\check{f}(\mathbf{1}_S) = f(S)$).
  2. that $\check{f}$ is convex (and consequently, that any arbitrary set function has a tight convex extension).
  3. that the convex closure $\check{f}$ is the convex envelope of the function defined only on the hypercube vertices, and that takes value $f(S)$ at $\mathbf{1}_S$.
  4. the definition of the Lovász extension of a set function, and that $\check{f}$ is the Lovász extension iff f is submodular.
- Note that the **concave** closure can also be defined, as $\hat{f}(x) = \max_{p \in \triangle^n(x)} E_{S \sim p}[f(S)]$, but it is in general impossible to obtain exactly even for submodular functions.

Greedy-based continuous extension of submodular f

- Given a submodular function f, a $w \in \cancel{\mathbb{R}^E_+}$ $[0,1]^E$, choose element order $(e_1, e_2, \dots, e_m)$ based on decreasing w,so that $1 \geq w(e_1) \geq w(e_2) \geq \dots \geq w(e_m) \geq 0$.
- Define chain $\emptyset \triangleq E_0 \subset E_1 \subset E_2 \subset \dots \subset E_m = E$ based on w, so the i^{th} element of this change has $E_i = \{e_1, e_2, \dots, e_i\}$.

We have, for $w \in \cancel{\mathbb{R}^E_+}$ $[0,1]^E$ that

$$\breve{f}(w) = \max(wx : x \in P_f) = \max(wx : x \in B_f) \tag{17.12}$$

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

$$= \sum_{i=1}^{m} w(e_i)(f(E_i) - f(E_{i-1})) \tag{17.14}$$

$$= (1 - w(e_1))f(E_0) + w(e_m)f(E_m) + \sum_{i=1}^{m-1} (w(e_i) - w(e_{i+1}))f(E_i) \tag{17.15}$$

Greedy-based continuous extension of submodular f

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

$$\breve{f}(w) = \max(wx : x \in B_f) \tag{17.12}$$

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

$$\breve{f}(w) = w(e_m)f(E_m) + \sum_{i=1}^{m-1}(w(e_i) - w(e_{i+1}))f(E_i) \tag{17.13}$$

$$= \sum_{i=1}^{m} \lambda_i f(E_i) \tag{17.14}$$

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.

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

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

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

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

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

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

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

Summary: comparison of the two extension forms

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

$$\breve{f}(w) = \max(wx : x \in B_f) = \sum_{i=1}^{m} \lambda_i f(E_i) \qquad (17.24)$$

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

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

$$\breve{f}(w) = \sum_{i=1}^{m} \lambda_i f(E_i) \qquad (17.25)$$

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

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

Lovász Extension, Submodularity and Convexity

Theorem 17.2.6

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

Proof.

- We've already seen that if f is submodular, its extension can be written via Eqn.(??) due to the greedy algorithm, and therefore is also equivalent to $\breve{f}(w) = \max\{wx : x \in P_f\}$, and thus is convex.
- Conversely, suppose the Lovász extension $\breve{f}(w) = \sum_i \lambda_i f(E_i)$ (of some function $f : 2^E \to \mathbb{R}$) is a convex function.
- We note that, based on the extension definition, in particular the definition of the $\{\lambda_i\}_i$, we have that $\breve{f}(\alpha w) = \alpha \breve{f}(w)$ for any $\alpha \in \mathbb{R}_+$. I.e., f is a positively homogeneous convex function.

...

Lovász ext. vs. the concave closure of submodular function

Theorem 17.2.6

Let $\breve{f}(w) = \max(wy : y \in B_f) = \sum_{i=1}^{m} \lambda_i f(E_i)$ *be the Lovász extension and* $\check{f}(x) = \min_{p \in \triangle^n(x)} E_{S \sim p}[f(S)]$ *be the convex closure. Then* $\breve{f}$ *and* $\check{f}$ *coincide iff* f *is submodular, i.e.,* $\breve{f}(w) = \check{f}(w), \forall w \in [0,1]$.

Proof.

- Assume f is submodular.
- Given x, let p^x be an achieving argmin in $\check{f}(x)$ that also maximizes $\sum_S p_S^x |S|^2$.
- Suppose $\exists A, B \subseteq V$ that are crossing (i.e., $A \not\subseteq B$, $B \not\subseteq A$) with positive p_A^x, p_B^x. W.l.o.g., $p_A^x \geq p_B^x > 0$.
- Then we may update p^x, keeping it a distribution, as follows:

$$\bar{p}_A^x \leftarrow p_A^x - p_B^x \qquad \bar{p}_B^x \leftarrow p_B^x - p_B^x \tag{17.33}$$

$$\bar{p}_{A\cup B}^x \leftarrow p_{A\cup B}^x + p_B^x \qquad \bar{p}_{A\cap B}^x \leftarrow p_{A\cap B}^x + p_B^x \tag{17.34}$$

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

...

Integration

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

Integration

- Integration is just summation (e.g., the $\int$ symbol has as its origins a sum).
- Lebesgue integration (See Rudin 1987, Definition 1.23, for details) allows integration w.r.t. an underlying measure μ of sets. Formal definition (from Rudin): given measurable function f, we can define

$$\int_X f du = \sup I_X(s) \tag{17.1}$$

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 $c_i > 0$.

Review

- Recall, a Boolean function f is any function $f : \{0,1\}^m \rightarrow \{0,1\}$ and is a pseudo-Boolean function if $f : \{0,1\}^m \rightarrow \mathbb{R}$.

Review

- Recall, 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 $x = \mathbf{1}_A$.

Review

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

Review

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

Choquet integral

Definition 17.3.1

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

$$C_f(w) = \sum_{i=1}^{m} (w_{e_i} - w_{e_{i+1}}) f(E_i) \tag{17.2}$$

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

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

$$C_f(w) = \sum_{i=1}^{m} w(e_i)(f(E_i) - f(E_{i-1})) \tag{17.3}$$

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

Choquet integral

Definition 17.3.1

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

$$C_f(w) = \sum_{i=1}^{m} (w_{e_i} - w_{e_{i+1}}) f(E_i) \tag{17.2}$$

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

$$\sum_{k=m}^{n} a_k b_k = \sum_{k=m}^{n} A_k(b_k - b_{k+1}) + A_n b_{n+1} - A_{m-1} b_m \tag{17.4}$$

The "integral" in the Choquet integral

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

The "integral" in the Choquet integral

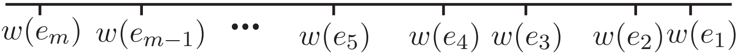
- Thought of as an integral over $\mathbb{R}$ of a piece-wise constant function.
- First note, assuming E is ordered according to descending w, so that $w(e_1) \geq w(e_2) \geq \cdots \geq w(e_{m-1}) \geq w(e_m)$, then $E_i = \{e_1, e_2, \ldots, e_i\} = \{e \in E : w_e \geq w_{e_i}\}$.

The "integral" in the Choquet integral

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

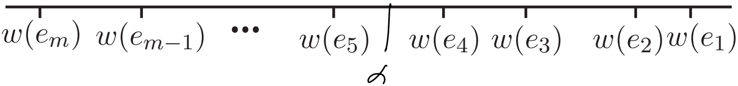
The "integral" in the Choquet integral

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



The "integral" in the Choquet integral

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



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

$$F_i(\alpha) = f(\{e \in E : w_e > \alpha\}) = f(E_i) \tag{17.5}$$

The "integral" in the Choquet integral

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

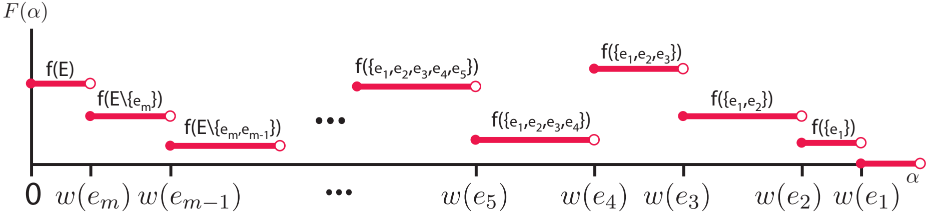
$$F(\alpha) = \begin{cases} f(E) & \text{if } 0 \leq \alpha < w_m \\ f(\{e \in E : w_e > \alpha\}) & \text{if } w_{e_{i+1}} \leq \alpha < w_{e_i},\ i \in \{1, \dots, m-1\} \\ 0\ (= f(\emptyset)) & \text{if } w_1 < \alpha \end{cases}$$

The "integral" in the Choquet integral

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

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

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



The "integral" in the Choquet integral

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

$$\tilde{f}(w) \stackrel{\text{def}}{=} \int_0^\infty F(\alpha) d\alpha \tag{17.6}$$

The "integral" in the Choquet integral

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

$$\tilde{f}(w) \stackrel{\text{def}}{=} \int_0^\infty F(\alpha) d\alpha \tag{17.6}$$

$$= \int_0^\infty f(\{e \in E : w_e > \alpha\}) d\alpha \tag{17.7}$$

The "integral" in the Choquet integral

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

$$\tilde{f}(w) \stackrel{\text{def}}{=} \int_0^\infty F(\alpha) d\alpha \tag{17.6}$$

$$= \int_0^\infty f(\{e \in E : w_e > \alpha\}) d\alpha \tag{17.7}$$

$$= \int_{w_{m+1}}^\infty f(\{e \in E : w_e > \alpha\}) d\alpha \tag{17.8}$$

The "integral" in the Choquet integral

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

$$\tilde{f}(w) \stackrel{\text{def}}{=} \int_0^\infty F(\alpha) d\alpha \tag{17.6}$$

$$= \int_0^\infty f(\{e \in E : w_e > \alpha\}) d\alpha \tag{17.7}$$

$$= \int_{w_{m+1}}^\infty f(\{e \in E : w_e > \alpha\}) d\alpha \tag{17.8}$$

$$= \sum_{i=1}^m \int_{w_{i+1}}^{w_i} f(\{e \in E : w_e > \alpha\}) d\alpha \tag{17.9}$$

The "integral" in the Choquet integral

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

$$\tilde{f}(w) \stackrel{\text{def}}{=} \int_0^\infty F(\alpha) d\alpha \tag{17.6}$$

$$= \int_0^\infty f(\{e \in E : w_e > \alpha\}) d\alpha \tag{17.7}$$

$$= \int_{w_{m+1}}^\infty f(\{e \in E : w_e > \alpha\}) d\alpha \tag{17.8}$$

$$= \sum_{i=1}^m \int_{w_{i+1}}^{w_i} f(\{e \in E : w_e > \alpha\}) d\alpha \tag{17.9}$$

$$= \sum_{i=1}^m \int_{w_{i+1}}^{w_i} f(E_i) d\alpha = \sum_{i=1}^m f(E_i)(w_i - w_{i+1}) \tag{17.10}$$

The "integral" in the Choquet integral

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

The "integral" in the Choquet integral

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

Definition 17.3.2

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

$$\tilde{f}(w) \stackrel{\text{def}}{=} \int_0^\infty F(\alpha) d\alpha \tag{17.11}$$

where the function F is defined as before.

The "integral" in the Choquet integral

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

Definition 17.3.2

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

$$\tilde{f}(w) \stackrel{\text{def}}{=} \int_0^\infty F(\alpha) d\alpha \tag{17.11}$$

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 17.3.2

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

$$\tilde{f}(w) \stackrel{\text{def}}{=} \int_0^\infty F(\alpha) d\alpha \tag{17.11}$$

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.

Integration, Aggregation, and Weighted Averages

- In finite discrete spaces, Lebesgue integration is just a weighted average, and can be seen as an aggregation function.

mean, max, min,
harmonic mean, geometric mean,
...

weighted avg,

Integration, Aggregation, and Weighted Averages

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

$$\text{WAVG}(x) = \sum_{e \in E} x(e) w(e) \tag{17.12}$$

Integration, Aggregation, and Weighted Averages

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

$$\text{WAVG}(x) = \sum_{e \in E} x(e)w(e) \tag{17.12}$$

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

$$\text{WAVG}(\mathbf{1}_e) = w(e) \tag{17.13}$$

Integration, Aggregation, and Weighted Averages

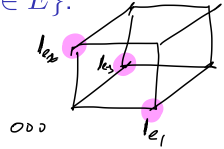
- In finite discrete spaces, Lebesgue integration is just a weighted average, and can be seen as an aggregation function.
- I.e., given a weight vector $w \in [0,1]^E$ for some finite ground set E, then for any $x \in \mathbb{R}^E$ we have the weighted average of x as:

$$\mathsf{WAVG}(x) = \sum_{e \in E} x(e)w(e) \tag{17.12}$$

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

$$\mathsf{WAVG}(\mathbf{1}_e) = w(e) \tag{17.13}$$

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



Integration, Aggregation, and Weighted Averages

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

$$\text{WAVG}(x) = \sum_{e \in E} x(e)w(e) \tag{17.12}$$

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

$$\text{WAVG}(\mathbf{1}_e) = w(e) \tag{17.13}$$

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

$$\text{WAVG}(x) = \sum_{e \in E} x(e)w(e) = \sum_{e \in E} x(e)\text{WAVG}(\mathbf{1}_e) \tag{17.14}$$

Integration, Aggregation, and Weighted Averages

$$\mathsf{WAVG}(x) = \sum_{e \in E} x(e) w(e) \tag{17.15}$$

- WAVG function is linear in weights w **and** 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 $\alpha \in \mathbb{R}$,

$$\mathsf{WAVG}_{w_1+w_2}(x) = \mathsf{WAVG}_{w_1}(x) + \mathsf{WAVG}_{w_2}(x), \tag{17.16}$$

$$\mathsf{WAVG}_{w}(x_1 + x_2) = \mathsf{WAVG}_{w}(x_1) + \mathsf{WAVG}_{w}(x_2), \tag{17.17}$$

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

$$\mathsf{WAVG}(\alpha x) = \alpha \mathsf{WAVG}(x). \tag{17.18}$$

Integration, Aggregation, and Weighted Averages

$LE_f(x)$

$LE_{f_1+f_2}(x) = LE_{f_1}(x) + LE_{f_2}(x)$

$$\text{WAVG}(x) = \sum_{e \in E} x(e)w(e) \tag{17.15}$$

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

$$\text{WAVG}_{w_1+w_2}(x) = \text{WAVG}_{w_1}(x) + \text{WAVG}_{w_2}(x), \tag{17.16}$$

$$\text{WAVG}_{w}(x_1+x_2) = \text{WAVG}_{w}(x_1) + \text{WAVG}_{w}(x_2), \tag{17.17}$$

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

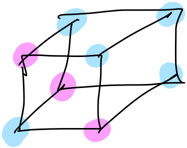
$$\text{WAVG}(\alpha x) = \alpha \text{WAVG}(x). \tag{17.18}$$

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

Integration, Aggregation, and Weighted Averages

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

$$\text{AG}(\mathbf{1}_A) = w_A \tag{17.19}$$



Integration, Aggregation, and Weighted Averages

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

$$\mathsf{AG}(\mathbf{1}_A) = w_A \tag{17.19}$$

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

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

$$= \sum_{e \in E} x(e) \sum_{A : A \ni e} w_A = \sum_{e \in E} x(e) w'_e$$

Choquet integral and aggregation

- We wish to produce some notion of generalized aggregation function having the flavor of:

$$\mathsf{AG}(x) = \sum_{A \subseteq E} x(A) w_A = \sum_{A \subseteq E} x(A) \mathsf{AG}(\mathbf{1}_A) \tag{17.21}$$

how does this correspond to Lovász extension?

Choquet integral and aggregation

- We wish to produce some notion of generalized aggregation function having the flavor of:

$$\mathsf{AG}(x) = \sum_{A \subseteq E} x(A) w_A = \sum_{A \subseteq E} x(A) \mathsf{AG}(\mathbf{1}_A) \tag{17.21}$$

  how does this correspond to Lovász extension?
- Let us partition the hypercube $[0,1]^m$ into q polytopes, $\mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_q$, each polytope defined by a set of vertices.

Choquet integral and aggregation

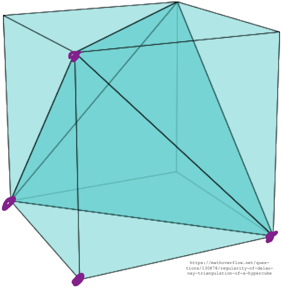
- We wish to produce some notion of generalized aggregation function having the flavor of:

$$\mathsf{AG}(x) = \sum_{A \subseteq E} x(A) w_A = \sum_{A \subseteq E} x(A) \mathsf{AG}(\mathbf{1}_A) \tag{17.21}$$

how does this correspond to Lovász extension?

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

- E.g., for each i, $\mathcal{V}_i = \{\mathbf{1}_{A_1}, \mathbf{1}_{A_2}, \dots, \mathbf{1}_{A_k}\}$ (k vertices) and the convex hull of $\mathcal{V}_i$ defines the i^{th} polytope. This forms a "triangulation" of the hypercube.



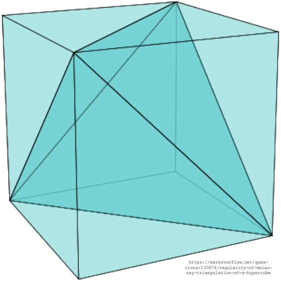
https://mathoverflow.net/questions/130878/regularity-of-delaunay-triangulation-of-a-hypercube

Choquet integral and aggregation

- We wish to produce some notion of generalized aggregation function having the flavor of:

$$\mathsf{AG}(x) = \sum_{A \subseteq E} x(A) w_A = \sum_{A \subseteq E} x(A) \mathsf{AG}(\mathbf{1}_A) \tag{17.21}$$

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



https://mathoverflow.net/questions/130878/regularity-of-delaunay-triangulation-of-a-hypercube

- For any $x \in [0,1]^m$ there is a (not necessarily unique) $\mathcal{V}(x) = \mathcal{V}_j$ for some j such that $x \in \text{conv}(\mathcal{V}(x))$.

Choquet integral and aggregation

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

$$\alpha_0^x(A) + \sum_{j=1}^{m} \alpha_j^x(A) x_j \in \mathbb{R} \tag{17.22}$$

Note that many of these coefficient are often zero.

Choquet integral and aggregation

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

$$\alpha_0^x(A) + \sum_{j=1}^m \alpha_j^x(A) x_j \in \mathbb{R} \tag{17.22}$$

  Note that many of these coefficient are often zero.
- From this, we can define an aggregation function of the form

$$\mathsf{AG}(x) \stackrel{\text{def}}{=} \sum_{A:\mathbf{1}_A \in \mathcal{V}(x)} \Big(\alpha_0^x(A) + \sum_{j=1}^m \alpha_j^x(A) x_j\Big) \mathsf{AG}(\mathbf{1}_A) \tag{17.23}$$

Choquet integral and aggregation

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

$$\alpha_0^x(A) + \sum_{j=1}^{m} \alpha_j^x(A) x_j \in \mathbb{R} \tag{17.22}$$

  Note that many of these coefficient are often zero.

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

$$\mathsf{AG}(x) \stackrel{\text{def}}{=} \sum_{A:\mathbf{1}_A \in \mathcal{V}(x)} \Big(\alpha_0^x(A) + \sum_{j=1}^{m} \alpha_j^x(A) x_j\Big) \mathsf{AG}(\mathbf{1}_A) \tag{17.23}$$

- Note, no longer necessarily linear in x.

Choquet integral and aggregation

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

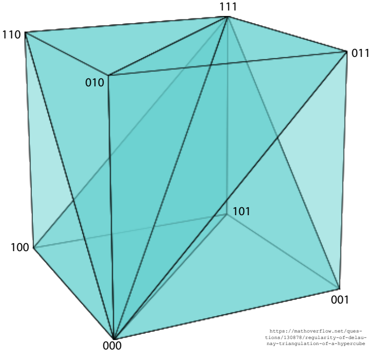
$$\text{conv}(\mathcal{V}_\sigma) = \{x \in [0,1]^n | x_{\sigma(1)} \geq x_{\sigma(2)} \geq \cdots \geq x_{\sigma(m)}\} \qquad (17.24)$$

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

$m = 3$

$m! = 6$

$\therefore$ 6 blocks in the partition



Choquet integral and aggregation

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

$$\mathrm{conv}(\mathcal{V}_\sigma) = \left\{x \in [0,1]^n | x_{\sigma(1)} \geq x_{\sigma(2)} \geq \cdots \geq x_{\sigma(m)}\right\} \qquad (17.24)$$

  Then these $m!$ blocks of the partition are called the canonical partitions of the hypercube.
- With this, we can define $\{\mathcal{V}_i\}_{i=1}^{m!}$ as the vertices of $\mathrm{conv}(\mathcal{V}_\sigma)$ for each permutation σ.

Choquet integral and aggregation

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

$$\operatorname{conv}(\mathcal{V}_\sigma) = \left\{x \in [0,1]^n | x_{\sigma(1)} \geq x_{\sigma(2)} \geq \cdots \geq x_{\sigma(m)}\right\} \tag{17.24}$$

  Then these $m!$ blocks of the partition are called the canonical partitions of the hypercube.
- With this, we can define $\{\mathcal{V}_i\}_{i=1}^{m!}$ as the vertices of $\operatorname{conv}(\mathcal{V}_\sigma)$ for each permutation σ. In this case, we have:

Choquet integral and aggregation

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

$$\mathrm{conv}(\mathcal{V}_\sigma) = \left\{x \in [0,1]^n | x_{\sigma(1)} \geq x_{\sigma(2)} \geq \cdots \geq x_{\sigma(m)}\right\} \tag{17.24}$$

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

- With this, we can define $\{\mathcal{V}_i\}_{i=1}^{m!}$ as the vertices of $\mathrm{conv}(\mathcal{V}_\sigma)$ for each permutation σ. In this case, we have:

Proposition 17.4.1

The above linear interpolation in Eqn. (17.23) using the canonical partition yields the Lovász extension with $\alpha_0^x(A) + \sum_{j=1}^m \alpha_j^x(A) x_j = x_{\sigma_i} - x_{\sigma_{i-1}}$ for $A = E_i = \{e_{\sigma_1}, \ldots, e_{\sigma_i}\}$ for appropriate order σ.

Choquet integral and aggregation

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

$$\operatorname{conv}(\mathcal{V}_\sigma) = \left\{x \in [0,1]^n | x_{\sigma(1)} \geq x_{\sigma(2)} \geq \cdots \geq x_{\sigma(m)}\right\} \tag{17.24}$$

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

- With this, we can define $\{\mathcal{V}_i\}_{i=1}^{m!}$ as the vertices of $\operatorname{conv}(\mathcal{V}_\sigma)$ for each permutation σ. In this case, we have:

Proposition 17.4.1

The above linear interpolation in Eqn. (17.23) using the canonical partition yields the Lovász extension with $\alpha_0^x(A) + \sum_{j=1}^m \alpha_j^x(A) x_j = x_{\sigma_i} - x_{\sigma_{i-1}}$ for $A = E_i = \{e_{\sigma_1}, \ldots, e_{\sigma_i}\}$ for appropriate order σ.

- Hence, Lovász extension can be seen as a form of generalized aggregation function. non-linear measure theory.

Lovász extension as max over orders

- We can also write the Lovász extension as follows:

$$\tilde{f}(w) = \max_{\sigma \in \Pi_{[m]}} w^{\intercal} c^{\sigma} \tag{17.25}$$

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

$$c_i^{\sigma} = f(E_{\sigma_i}) - f(E_{\sigma_{i-1}}) \tag{17.26}$$

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

Lovász extension as max over orders

- We can also write the Lovász extension as follows:

$$\tilde{f}(w) = \max_{\sigma \in \Pi_{[m]}} w^\intercal c^\sigma \tag{17.25}$$

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

$$c_i^\sigma = f(E_{\sigma_i}) - f(E_{\sigma_{i-1}}) = f(e_{\sigma_i} / E_{\sigma_{i-1}}) \tag{17.26}$$

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

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

$$\tilde{f}(w) = \max_{x \in P_f} w^\intercal x = \max_{x \in B_f} w^\intercal x \tag{17.27}$$

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

Lovász extension, defined in multiple ways

- As shorthand notation, lets use $\{w \geq \alpha\} \equiv \{e \in E : w(e) \geq \alpha\}$, called the (weak) α-superlevel set of w.

Lovász extension, defined in multiple ways

- As shorthand notation, lets use $\{w \geq \alpha\} \equiv \{e \in E : w(e) \geq \alpha\}$, called the (weak) α-superlevel set of w. A similar definition holds for $\{w > \alpha\}$ (called the strong, or strict, α-superlevel set of w).

Lovász extension, defined in multiple ways

- As shorthand notation, lets use $\{w \geq \alpha\} \equiv \{e \in E : w(e) \geq \alpha\}$, called the (weak) α-superlevel set of w. A similar definition holds for $\{w > \alpha\}$ (called the strong, or strict, α-superlevel set of w).
- Given any $w \in \mathbb{R}^E$, sort E as $w(e_1) \geq w(e_2) \geq \cdots \geq w(e_m)$.

Lovász extension, defined in multiple ways

- As shorthand notation, lets use $\{w \geq \alpha\} \equiv \{e \in E : w(e) \geq \alpha\}$, called the (weak) α-superlevel set of w. A similar definition holds for $\{w > \alpha\}$ (called the strong, or strict, α-superlevel set of w).
- Given any $w \in \mathbb{R}^E$, sort E as $w(e_1) \geq w(e_2) \geq \cdots \geq w(e_m)$. Also, w.l.o.g., number elements of w so that $w_1 \geq w_2 \geq \cdots \geq w_m$.

Lovász extension, defined in multiple ways

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

$$\tilde{f}(w) = \sum_{i=1}^{m} w(e_i) f(e_i | E_{i-1}) \tag{17.28}$$

$$= \sum_{i=1}^{m-1} f(E_i)(w(e_i) - w(e_{i+1})) + f(E)w(e_m) \tag{17.29}$$

$$= \sum_{i=1}^{m-1} \lambda_i f(E_i) \tag{17.30}$$

Lovász extension, as integral

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

$$\tilde{f}(w) = \sum_{i=1}^{m} w(e_i) f(e_i | E_{i-1}) = \sum_{i=1}^{m} \lambda_i f(E_i) \tag{17.31}$$

$$= \sum_{i=1}^{m-1} f(E_i)(w(e_i) - w(e_{i+1})) + f(E) w(e_m) \tag{17.32}$$

$$= \int_{\min\{w_1,\dots,w_m\}}^{+\infty} f(\{w \geq \alpha\}) d\alpha + f(E) \min\{w_1, \dots, w_m\} \tag{17.33}$$

$$\stackrel{\text{(a)}}{=} \int_0^{+\infty} f(\{w \geq \alpha\}) d\alpha + \int_{-\infty}^{0} [f(\{w \geq \alpha\}) - f(E)] d\alpha \tag{17.34}$$

We will show (a) in a few slides.

general Lovász extension, as simple integral

- Assuming (a), we have that, given function f, and any $w \in \mathbb{R}^E$:

$$\tilde{f}(w) = \int_{-\infty}^{+\infty} \hat{f}(\alpha) d\alpha \tag{17.35}$$

where

$$\hat{f}(\alpha) = \begin{cases} f(\{w \geq \alpha\}) & \text{if } \alpha \geq 0 \\ f(\{w \geq \alpha\}) - f(E) & \text{if } \alpha < 0 \end{cases} \tag{17.36}$$

general Lovász extension, as simple integral

- Assuming (a), we have that, given function f, and any $w \in \mathbb{R}^E$:

$$\tilde{f}(w) = \int_{-\infty}^{+\infty} \hat{f}(\alpha) d\alpha \tag{17.35}$$

where

$$\hat{f}(\alpha) = \begin{cases} f(\{w \geq \alpha\}) & \text{if } \alpha \geq 0 \\ f(\{w \geq \alpha\}) - f(E) & \text{if } \alpha < 0 \end{cases} \tag{17.36}$$

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

general Lovász extension, as simple integral

- Assuming (a), we have that, given function f, and any $w \in \mathbb{R}^E$:

$$\tilde{f}(w) = \int_{-\infty}^{+\infty} \hat{f}(\alpha) d\alpha \tag{17.35}$$

where

$$\hat{f}(\alpha) = \begin{cases} f(\{w \geq \alpha\}) & \text{if } \alpha \geq 0 \\ f(\{w \geq \alpha\}) - f(E) & \text{if } \alpha < 0 \end{cases} \tag{17.36}$$

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

Lovász extension, as integral

- To show Eqn. (17.33), first note that the r.h.s. terms are the same since $w(e_m) = \min\{w_1, \dots, w_m\}$.

Lovász extension, as integral

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

$$\tilde{f}(w) = \sum_{i=1}^{m} w(e_i) f(e_i|E_{i-1}) = \sum_{i=1}^{m} \lambda_i f(E_i) \tag{17.31}$$

$$= \sum_{i=1}^{m-1} f(E_i)(w(e_i) - w(e_{i+1})) + f(E) w(e_m) \tag{17.32}$$

$$= \int_{\min\{w_1,\dots,w_m\}}^{+\infty} f(\{w \geq \alpha\}) d\alpha + f(E) \min\{w_1, \dots, w_m\} \tag{17.33}$$

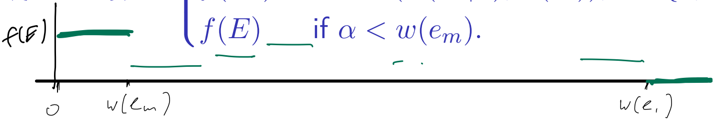
$$\overset{(a)}{=} \int_{0}^{+\infty} f(\{w \geq \alpha\}) d\alpha + \int_{-\infty}^{0} [f(\{w \geq \alpha\}) - f(E)] d\alpha \tag{17.34}$$

We will show (a) in a few slides.

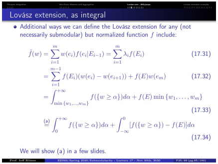
Lovász extension, as integral

- To show Eqn. (17.33), first note that the r.h.s. terms are the same since $w(e_m) = \min\{w_1, \dots, w_m\}$.
- Then, consider that, as a function of α, we have

$$f(\{w \geq \alpha\}) = \begin{cases} 0 & \text{if } \alpha > w(e_1) \\ f(E_k) & \text{if } \alpha \in (w(e_{k+1}), w(e_k)), k \in \{1, \dots, m-1\} \\ f(E) & \text{if } \alpha < w(e_m). \end{cases} \tag{17.37}$$



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



Lovász extension, as integral

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

$$\tilde{f}(w) = \sum_{i=1}^{m} w(e_i) f(e_i | E_{i-1}) = \sum_{i=1}^{m} \lambda_i f(E_i) \tag{17.31}$$

$$= \sum_{i=1}^{m-1} f(E_i)(w(e_i) - w(e_{i+1})) + f(E) w(e_m) \tag{17.32}$$

$$= \int_{\min\{w_1, \dots, w_m\}}^{+\infty} f(\{w \geq \alpha\}) d\alpha + f(E) \min\{w_1, \dots, w_m\} \tag{17.33}$$

$$\stackrel{(a)}{=} \int_0^{+\infty} f(\{w \geq \alpha\}) d\alpha + \int_{-\infty}^{0} [f(\{w \geq \alpha\}) - f(E)] d\alpha \tag{17.34}$$

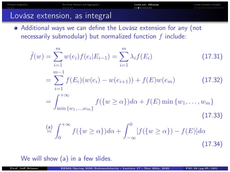
We will show (a) in a few slides.

Lovász extension, as integral

- To show Eqn. (17.33), first note that the r.h.s. terms are the same since $w(e_m) = \min\{w_1, \ldots, w_m\}$.
- Then, consider that, as a function of α, we have

$$f(\{w \geq \alpha\}) = \begin{cases} 0 & \text{if } \alpha > w(e_1) \\ f(E_k) & \text{if } \alpha \in (w(e_{k+1}), w(e_k)), k \in \{1, \ldots, m-1\} \\ f(E) & \text{if } \alpha < w(e_m). \end{cases} \tag{17.37}$$

  For integration purposes, we may use open intervals since sets of zero measure don't change integration.
- Inside the integral, then, this recovers Eqn. (17.32).


Lovász extension, as integral

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

$$\tilde{f}(w) = \sum_{i=1}^{m} w(e_i) f(e_i | E_{i-1}) = \sum_{i=1}^{m} \lambda_i f(E_i) \tag{17.31}$$

$$= \sum_{i=1}^{m-1} f(E_i)(w(e_i) - w(e_{i+1})) + f(E)w(e_m) \tag{17.32}$$

$$= \int_{\min\{w_1,\ldots,w_m\}}^{+\infty} f(\{w \geq \alpha\}) d\alpha + f(E) \min\{w_1, \ldots, w_m\} \tag{17.33}$$

$$\stackrel{(a)}{=} \int_0^{+\infty} f(\{w \geq \alpha\}) d\alpha + \int_{-\infty}^{0} [f(\{w \geq \alpha\}) - f(E)] d\alpha \tag{17.34}$$

We will show (a) in a few slides.

Prof. Jeff Bilmes EE563/Spring 2020/Submodularity - Lecture 17 - Nov 30th, 2020 F25/49 (pg.65/160)

Lovász extension, as integral

- To show Eqn. (17.34), start with Eqn. (17.33), note $w_m = \min\{w_1, \dots, w_m\}$, take any $\beta \leq \min\{0, w_1, \dots, w_m\}$, and form:

$\tilde{f}(w)$ $\therefore \beta \leq \min(0, w_m)$

Choquet Integration | Non-linear Measure and Aggregation | Lovász extn., defs/props | Lovász extension examples

Lovász extension, as integral

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

$$\tilde{f}(w) = \sum_{i=1}^{m} w(e_i) f(e_i | E_{i-1}) = \sum_{i=1}^{m} \lambda_i f(E_i) \tag{17.31}$$

$$= \sum_{i=1}^{m-1} f(E_i)(w(e_i) - w(e_{i+1})) + f(E) w(e_m) \tag{17.32}$$

$$= \int_{\min\{w_1,\dots,w_m\}}^{+\infty} f(\{w \geq \alpha\}) d\alpha + f(E) \min\{w_1, \dots, w_m\} \tag{17.33}$$

$$\overset{(a)}{=} \int_{0}^{+\infty} f(\{w \geq \alpha\}) d\alpha + \int_{-\infty}^{0} [f(\{w \geq \alpha\}) - f(E)] d\alpha \tag{17.34}$$

We will show (a) in a few slides.

Lovász extension, as integral

- To show Eqn. (17.34), start with Eqn. (17.33), note $w_m = \min\{w_1, \dots, w_m\}$, take any $\beta \leq \min\{0, w_1, \dots, w_m\}$, and form:
$$\tilde{f}(w) = \int_{w_m}^{+\infty} f(\{w \geq \alpha\})d\alpha + f(E)\min\{w_1, \dots, w_m\}$$

Lovász extension, as integral

- To show Eqn. (17.34), start with Eqn. (17.33), note $w_m = \min\{w_1, \ldots, w_m\}$, take any $\beta \le \min\{0, w_1, \ldots, w_m\}$, and form:

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

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

Lovász extension, as integral

- To show Eqn. (17.34), start with Eqn. (17.33), note $w_m = \min\{w_1, \ldots, w_m\}$, take any $\beta \leq \min\{0, w_1, \ldots, w_m\}$, and form:

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

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

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

Lovász extension, as integral

- To show Eqn. (17.34), start with Eqn. (17.33), note $w_m = \min\{w_1, \ldots, w_m\}$, take any $\beta \leq \min\{0, w_1, \ldots, w_m\}$, and form:

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

Lovász extension, as integral

- To show Eqn. (17.34), start with Eqn. (17.33), note $w_m = \min\{w_1, \dots, w_m\}$, take any $\beta \le \min\{0, w_1, \dots, w_m\}$, and form:

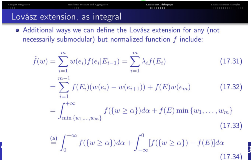
$$\tilde{f}(w) = \int_{w_m}^{+\infty} f(\{w \ge \alpha\})d\alpha + f(E)\min\{w_1, \dots, w_m\}$$

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

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

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

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



Lovász extension, as integral

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

$$\tilde{f}(w) = \sum_{i=1}^{m} w(e_i)f(e_i|E_{i-1}) = \sum_{i=1}^{m} \lambda_i f(E_i) \tag{17.31}$$

$$= \sum_{i=1}^{m-1} f(E_i)(w(e_i) - w(e_{i+1})) + f(E)w(e_m) \tag{17.32}$$

$$= \int_{\min\{w_1,\dots,w_m\}}^{+\infty} f(\{w \ge \alpha\})d\alpha + f(E)\min\{w_1, \dots, w_m\} \tag{17.33}$$

$$\overset{(a)}{=} \int_0^{+\infty} f(\{w \ge \alpha\})d\alpha + \int_{-\infty}^{0} [f(\{w \ge \alpha\}) - f(E)]d\alpha \tag{17.34}$$

Lovász extension, as integral

- To show Eqn. (17.34), start with Eqn. (17.33), note $w_m = \min\{w_1, \dots, w_m\}$, take any $\beta \le \min\{0, w_1, \dots, w_m\}$, and form:

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

and then let $\beta \to -\infty$ and we get Eqn. (17.34), i.e.:

Lovász extension, as integral

- To show Eqn. (17.34), start with Eqn. (17.33), note $w_m = \min\{w_1, \ldots, w_m\}$, take any $\beta \leq \min\{0, w_1, \ldots, w_m\}$, and form:

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

and then let $\beta \to -\infty$ and we get Eqn. (17.34), i.e.:

$$= \int_{0}^{+\infty} f(\{w \geq \alpha\})d\alpha + \int_{-\infty}^{0} [f(\{w \geq \alpha\}) - f(E)]d\alpha \qquad (17.34)$$

Lovász extension properties

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

Lovász extension properties

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

Theorem 17.5.1

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

Lovász extension properties

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

Theorem 17.5.1

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

1. *Superposition of LE operator: Given f and g with Lovász extensions $\tilde{f}$ and $\tilde{g}$ then $\tilde{f} + \tilde{g}$ is the Lovász extension of $f + g$ and $\lambda \tilde{f}$ is the Lovász extension of λf for $\lambda \in \mathbb{R}$.*

$$\cancel{LE(\lambda x) = \lambda LE(x)}$$

$$LE_{\lambda f}(x) = \lambda \cdot LE_f(x)$$

Lovász extension properties

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

Theorem 17.5.1

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

1. *Superposition of LE operator: Given f and g with Lovász extensions $\tilde{f}$ and $\tilde{g}$ then $\tilde{f} + \tilde{g}$ is the Lovász extension of $f + g$ and $\lambda \tilde{f}$ is the Lovász extension of λf for $\lambda \in \mathbb{R}$.*
2. *If $w \in \mathbb{R}_+^E$ then $\tilde{f}(w) = \int_0^{+\infty} f(\{w \geq \alpha\}) d\alpha$.*

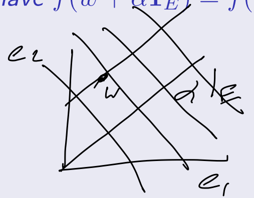
Lovász extension properties

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

Theorem 17.5.1

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

1. *Superposition of LE operator: Given f and g with Lovász extensions $\tilde{f}$ and $\tilde{g}$ then $\tilde{f} + \tilde{g}$ is the Lovász extension of $f + g$ and $\lambda \tilde{f}$ is the Lovász extension of λf for $\lambda \in \mathbb{R}$.*
2. *If $w \in \mathbb{R}_+^E$ then $\tilde{f}(w) = \int_0^{+\infty} f(\{w \geq \alpha\}) d\alpha$.*
3. *For $w \in \mathbb{R}^E$, and $\alpha \in \mathbb{R}$, we have $\tilde{f}(w + \alpha \mathbf{1}_E) = \tilde{f}(w) + \alpha f(E)$.*



Lovász extension properties

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

Theorem 17.5.1

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

1. *Superposition of LE operator: Given f and g with Lovász extensions $\tilde{f}$ and $\tilde{g}$ then $\tilde{f} + \tilde{g}$ is the Lovász extension of $f + g$ and $\lambda \tilde{f}$ is the Lovász extension of λf for $\lambda \in \mathbb{R}$.*
2. *If $w \in \mathbb{R}_+^E$ then $\tilde{f}(w) = \int_0^{+\infty} f(\{w \geq \alpha\}) d\alpha$.*
3. *For $w \in \mathbb{R}^E$, and $\alpha \in \mathbb{R}$, we have $\tilde{f}(w + \alpha \mathbf{1}_E) = \tilde{f}(w) + \alpha f(E)$.*
4. *Positive homogeneity: I.e., $\tilde{f}(\alpha w) = \alpha \tilde{f}(w)$ for $\alpha \geq 0$.*

Lovász extension properties

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

Theorem 17.5.1

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

1. *Superposition of LE operator: Given f and g with Lovász extensions $\tilde{f}$ and $\tilde{g}$ then $\tilde{f} + \tilde{g}$ is the Lovász extension of $f + g$ and $\lambda \tilde{f}$ is the Lovász extension of λf for $\lambda \in \mathbb{R}$.*
2. *If $w \in \mathbb{R}_+^E$ then $\tilde{f}(w) = \int_0^{+\infty} f(\{w \geq \alpha\}) d\alpha$.*
3. *For $w \in \mathbb{R}^E$, and $\alpha \in \mathbb{R}$, we have $\tilde{f}(w + \alpha \mathbf{1}_E) = \tilde{f}(w) + \alpha f(E)$.*
4. *Positive homogeneity: I.e., $\tilde{f}(\alpha w) = \alpha \tilde{f}(w)$ for $\alpha \geq 0$.*
5. *For all $A \subseteq E$, $\tilde{f}(\mathbf{1}_A) = f(A)$.*

Lovász extension properties

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

Theorem 17.5.1

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

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

$f(A) = f'(A) + f'(V \setminus A)$

Lovász extension properties

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

Theorem 17.5.1

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

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

Lovász extension properties: ex. property 3

- Consider property property 3, for example, which says that $\tilde{f}(w + \alpha \mathbf{1}_E) = \tilde{f}(w) + \alpha f(E)$.

Lovász extension properties: ex. property 3

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

Lovász extension properties: ex. property 3

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

Lovász extension properties

- Given Eqns. (17.31) through (17.34), most of the above properties are relatively easy to derive.
- For example, if f is symmetric, and since $f(E) = f(\emptyset) = 0$, we have

$$\tilde{f}(-w) = \int_{-\infty}^{\infty} f(\{-w \geq \alpha\})d\alpha \tag{17.40}$$

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. (17.31) through (17.34), most of the above properties are relatively easy to derive.
- For example, if f is symmetric, and since $f(E) = f(\emptyset) = 0$, we have

$$\tilde{f}(-w) = \int_{-\infty}^{\infty} f(\{-w \geq \alpha\})d\alpha = \int_{-\infty}^{\infty} f(\{w \leq -\alpha\})d\alpha \tag{17.38}$$

(17.40)

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. (17.31) through (17.34), most of the above properties are relatively easy to derive.
- For example, if f is symmetric, and since $f(E) = f(\emptyset) = 0$, we have

$$\tilde{f}(-w) = \int_{-\infty}^{\infty} f(\{-w \geq \alpha\})d\alpha = \int_{-\infty}^{\infty} f(\{w \leq -\alpha\})d\alpha \tag{17.38}$$

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

$$\tag{17.40}$$

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. (17.31) through (17.34), most of the above properties are relatively easy to derive.
- For example, if f is symmetric, and since $f(E) = f(\emptyset) = 0$, we have

$$\tilde{f}(-w) = \int_{-\infty}^{\infty} f(\{-w \geq \alpha\})d\alpha = \int_{-\infty}^{\infty} f(\{w \leq -\alpha\})d\alpha \tag{17.38}$$

$$\stackrel{(a)}{=} \int_{-\infty}^{\infty} f(\{w \leq \alpha\})d\alpha \stackrel{(b)}{=} \int_{-\infty}^{\infty} f(\{w > \alpha\})d\alpha \tag{17.39}$$

$$\{w > \alpha\} = E \setminus \{w \leq \alpha\} \tag{17.40}$$

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. (17.31) through (17.34), most of the above properties are relatively easy to derive.
- For example, if f is symmetric, and since $f(E) = f(\emptyset) = 0$, we have

$$\tilde{f}(-w) = \int_{-\infty}^{\infty} f(\{-w \geq \alpha\})d\alpha = \int_{-\infty}^{\infty} f(\{w \leq -\alpha\})d\alpha \tag{17.38}$$

$$\stackrel{(a)}{=} \int_{-\infty}^{\infty} f(\{w \leq \alpha\})d\alpha \stackrel{(b)}{=} \int_{-\infty}^{\infty} f(\{w > \alpha\})d\alpha \tag{17.39}$$

$$= \int_{-\infty}^{\infty} f(\{w \geq \alpha\})d\alpha \tag{17.40}$$

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. (17.31) through (17.34), most of the above properties are relatively easy to derive.
- For example, if f is symmetric, and since $f(E) = f(\emptyset) = 0$, we have

$$\tilde{f}(-w) = \int_{-\infty}^{\infty} f(\{-w \geq \alpha\}) d\alpha = \int_{-\infty}^{\infty} f(\{w \leq -\alpha\}) d\alpha \tag{17.38}$$

$$\stackrel{(a)}{=} \int_{-\infty}^{\infty} f(\{w \leq \alpha\}) d\alpha \stackrel{(b)}{=} \int_{-\infty}^{\infty} f(\{w > \alpha\}) d\alpha \tag{17.39}$$

$$= \int_{-\infty}^{\infty} f(\{w \geq \alpha\}) d\alpha = \tilde{f}(w) \tag{17.40}$$

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, expected value of random variable

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

Lovász extension, expected value of random variable

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

Lovász extension, expected value of random variable

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

Lovász extension, expected value of random variable

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

Lovász extension, expected value of random variable

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

$$\tilde{f}(w) = \mathbb{E}_{p(\alpha)}[\underbrace{f(\{w \geq \alpha\})}_{h(\alpha)}] = \mathbb{E}_{p(\alpha)}[\underbrace{f(e \in E : w(e_i) \geq \alpha)}_{h(\alpha)}] \quad (17.41)$$

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

Lovász extension, expected value of random variable

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

$$\tilde{f}(w) = \mathbb{E}_{p(\alpha)}[\underbrace{f(\{w \geq \alpha\})}_{h(\alpha)}] = \mathbb{E}_{p(\alpha)}[\underbrace{f(e \in E : w(e_i) \geq \alpha)}_{h(\alpha)}] \tag{17.41}$$

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

One slide review of ~~co~~ncave relaxation

[illegible]

- convex closure $\check{f}(x) = \min_{p \in \triangle^n(x)} E_{S \sim p}[f(S)]$, where where $\triangle^n(x) = \left\{ p \in \mathbb{R}^{2^n} : \sum_{S \subseteq V} p_S = 1,\ p_S \geq 0 \forall S \subseteq V, \ \&\ \sum_{S \subseteq V} p_S \mathbf{1}_S = x \right\}$
- "Edmonds" extension $\breve{f}(w) = \max(wx : x \in B_f)$
- Lovász extension $f_{\mathsf{LE}}(w) = \sum_{i=1}^m \lambda_i f(E_i)$, with λ_i such that $w = \sum_{i=1}^m \lambda_i \mathbf{1}_{E_i}$
- $\tilde{f}(w) = \max_{\sigma \in \Pi_{[m]}} w^\intercal c^\sigma$, $\Pi_{[m]}$ set of $m!$ permutations of $[m]$, $\sigma \in \Pi_{[m]}$ a permutation, c^σ vector with $c_i^\sigma = f(E_{\sigma_i}) - f(E_{\sigma_{i-1}})$, $E_{\sigma_i} = \{e_{\sigma_1}, e_{\sigma_2}, \ldots, e_{\sigma_i}\}$.
- Choquet integral $C_f(w) = \sum_{i=1}^m (w_{e_i} - w_{e_{i+1}}) f(E_i)$
- $\tilde{f}(w) = \int_{-\infty}^{+\infty} \hat{f}(\alpha) d\alpha$, $\hat{f}(\alpha) = \begin{cases} f(\{w \geq \alpha\}) & \text{if } \alpha \geq 0 \\ f(\{w \geq \alpha\}) - f(E) & \text{if } \alpha < 0 \end{cases}$
- All the same when f is submodular.

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

- If $w_1 \geq w_2$, then

$$\tilde{f}(w) = w_1 f(\{1\}) + w_2 f(\{2\}|\{1\}) \tag{17.42}$$

$$= (w_1 - w_2) f(\{1\}) + w_2 f(\{1, 2\}) \tag{17.43}$$

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

- If $w_1 \geq w_2$, then

$$\tilde{f}(w) = w_1 f(\{1\}) + w_2 f(\{2\}|\{1\}) \tag{17.42}$$
$$= (w_1 - w_2) f(\{1\}) + w_2 f(\{1, 2\}) \tag{17.43}$$

- If $w_1 \leq w_2$, then

$$\tilde{f}(w) = w_2 f(\{2\}) + w_1 f(\{1\}|\{2\}) \tag{17.44}$$
$$= (w_2 - w_1) f(\{2\}) + w_1 f(\{1, 2\}) \tag{17.45}$$

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

- If $w_1 \geq w_2$, then

$$\tilde{f}(w) = w_1 f(\{1\}) + w_2 f(\{2\}|\{1\}) \tag{17.46}$$

$$= (w_1 - w_2) f(\{1\}) + w_2 f(\{1,2\}) \tag{17.47}$$

$$= \frac{1}{2} f(1)(w_1 - w_2) + \frac{1}{2} f(1)(w_1 - w_2) \tag{17.48}$$

$$+ \frac{1}{2} f(\{1,2\})(w_1 + w_2) - \frac{1}{2} f(\{1,2\})(w_1 - w_2) \tag{17.49}$$

$$+ \frac{1}{2} f(2)(w_1 - w_2) + \frac{1}{2} f(2)(w_2 - w_1) \tag{17.50}$$

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

- If $w_1 \geq w_2$, then

$$\tilde{f}(w) = w_1 f(\{1\}) + w_2 f(\{2\}|\{1\}) \tag{17.46}$$
$$= (w_1 - w_2) f(\{1\}) + w_2 f(\{1,2\}) \tag{17.47}$$
$$= \frac{1}{2} f(1)(w_1 - w_2) + \frac{1}{2} f(1)(w_1 - w_2) \tag{17.48}$$
$$+ \frac{1}{2} f(\{1,2\})(w_1 + w_2) - \frac{1}{2} f(\{1,2\})(w_1 - w_2) \tag{17.49}$$
$$+ \frac{1}{2} f(2)(w_1 - w_2) + \frac{1}{2} f(2)(w_2 - w_1) \tag{17.50}$$

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

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

- This gives, for general w_1, w_2, that

$$\tilde{f}(w) = \frac{1}{2}\left(f(\{1\}) + f(\{2\}) - f(\{1,2\})\right)|w_1 - w_2| \tag{17.51}$$

$$+ \frac{1}{2}\left(f(\{1\}) - f(\{2\}) + f(\{1,2\})\right) w_1 \tag{17.52}$$

$$+ \frac{1}{2}\left(-f(\{1\}) + f(\{2\}) + f(\{1,2\})\right) w_2 \tag{17.53}$$

$$= -\left(f(\{1\}) + f(\{2\}) - f(\{1,2\})\right) \min\{w_1, w_2\} \tag{17.54}$$

$$+ f(\{1\})w_1 + f(\{2\})w_2 \tag{17.55}$$

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

- This gives, for general w_1, w_2, that

$$\tilde{f}(w) = \frac{1}{2}\left(f(\{1\}) + f(\{2\}) - f(\{1,2\})\right)|w_1 - w_2| \tag{17.51}$$

$$+ \frac{1}{2}\left(f(\{1\}) - f(\{2\}) + f(\{1,2\})\right) w_1 \tag{17.52}$$

$$+ \frac{1}{2}\left(-f(\{1\}) + f(\{2\}) + f(\{1,2\})\right) w_2 \tag{17.53}$$

$$= -\left(f(\{1\}) + f(\{2\}) - f(\{1,2\})\right) \min\{w_1, w_2\} \tag{17.54}$$

$$+ f(\{1\})w_1 + f(\{2\})w_2 \tag{17.55}$$

- 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.

Combinatorial MI.

$I_f(e_1; e_2) = f(e_1) + f(e_2) - f(e_1, e_2)$

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

- This gives, for general w_1, w_2, that

$$\begin{aligned}
\tilde{f}(w) &= \frac{1}{2}\left(f(\{1\}) + f(\{2\}) - f(\{1,2\})\right)|w_1 - w_2| && (17.51)\\
&\quad + \frac{1}{2}\left(f(\{1\}) - f(\{2\}) + f(\{1,2\})\right) w_1 && (17.52)\\
&\quad + \frac{1}{2}\left(-f(\{1\}) + f(\{2\}) + f(\{1,2\})\right) w_2 && (17.53)\\
&= -\left(f(\{1\}) + f(\{2\}) - f(\{1,2\})\right) \min\{w_1, w_2\} && (17.54)\\
&\quad + f(\{1\})w_1 + f(\{2\})w_2 && (17.55)
\end{aligned}$$

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

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

- If $w_1 \geq w_2$, then

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

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

- If $w_1 \geq w_2$, then

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

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

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

- If $w_1 \geq w_2$, then

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

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

$$= \left(\frac{1}{f(1,2)}, \frac{1}{f(1,2)} \right)$$

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

- If $w_1 \geq w_2$, then

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

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

- If $w_1 \leq w_2$, then

$$\tilde{f}(w) = w_2 f(\{2\}) + w_1 f(\{1\}|\{2\}) \tag{17.57}$$

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

- If $w_1 \geq w_2$, then

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

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

- If $w_1 \leq w_2$, then

$$\tilde{f}(w) = w_2 f(\{2\}) + w_1 f(\{1\}|\{2\}) \tag{17.57}$$

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

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

- If $w_1 \geq w_2$, then

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

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

- If $w_1 \leq w_2$, then

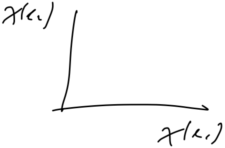
$$\tilde{f}(w) = w_2 f(\{2\}) + w_1 f(\{1\}|\{2\}) \tag{17.57}$$

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

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

- If $w_1 \geq w_2$, then

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



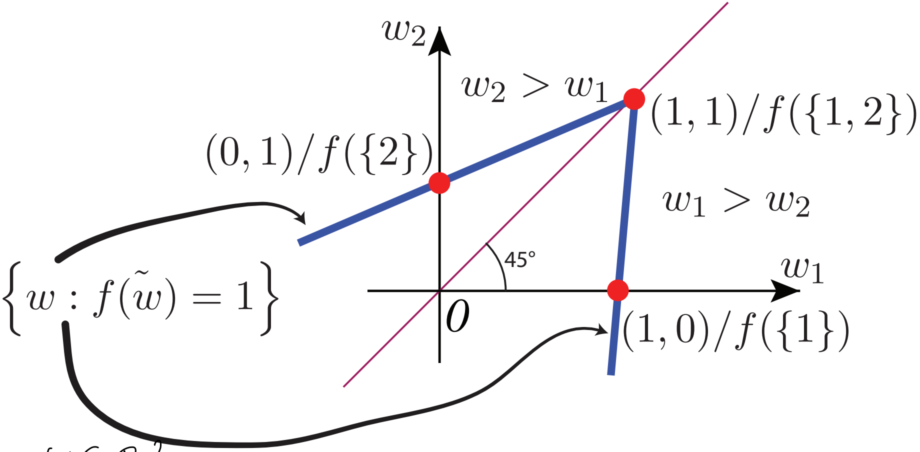
  - If $w = (1, 0)/f(\{1\}) = \Big(1/f(\{1\}), 0\Big)$ then $\tilde{f}(w) = 1$.
  - If $w = (1, 1)/f(\{1, 2\})$ then $\tilde{f}(w) = 1$.
- If $w_1 \leq w_2$, then

$$\tilde{f}(w) = w_2 f(\{2\}) + w_1 f(\{1\}|\{2\}) \tag{17.57}$$

  - If $w = (0, 1)/f(\{2\}) = (0, 1/f(\{2\}))$ then $\tilde{f}(w) = 1$.
  - If $w = (1, 1)/f(\{1, 2\})$ then $\tilde{f}(w) = 1$.
- Can plot contours of the form $\left\{w \in \mathbb{R}^2 : \tilde{f}(w) = 1\right\}$, particular marked points of form $w = \mathbf{1}_A \times \frac{1}{f(A)}$ for certain A, where $\tilde{f}(w) = 1$.

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

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



$w \in \mathbb{R}^2$

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 $f(E) = f'(E) - x(E) = 0$.

$x \in B_{f'}$, a vertex of $B_{f'}$ generated from greedy according to the order σ,

$$E_i = \{e_{\sigma_1}, e_{\sigma_2}, \dots, e_{\sigma_i}\}$$

$$\therefore \quad f(E_i) = 0 \quad \forall i.$$

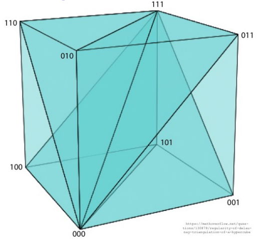
$$E_m = E$$

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

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

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

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

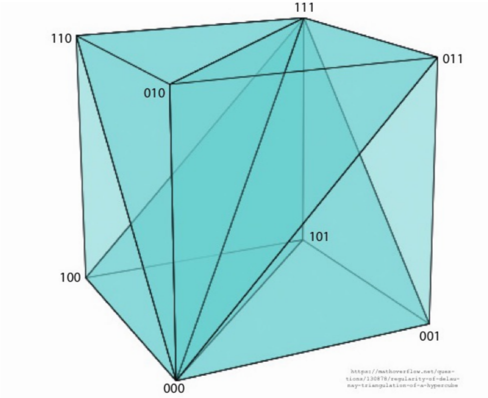


Example: $m = 3$, $E = \{1, 2, 3\}$

- In order to visualize in 3D, we make a few simplifications.
- Consider any submodular f' and $x \in B_{f'}$. Then $f(A) = f'(A) - x(A)$ is submodular, and moreover $f(E) = f'(E) - x(E) = 0$.
- Hence, from $\tilde{f}(w + \alpha \mathbf{1}_E) = \tilde{f}(w) + \alpha f(E)$, we have that $\tilde{f}(w + \alpha \mathbf{1}_E) = \tilde{f}(w)$ when $f(E) = 0$.
- Thus, we can look "down" on the contour plot of the Lovász extension, $\left\{ w : \tilde{f}(w) = 1 \right\}$, from a vantage point of being right on the ray $\{x : x = \alpha \mathbf{1}_E, \alpha > 0\}$ since moving in direction $\mathbf{1}_E$ changes nothing.
- I.e., consider 2D plane perpendicular to the ray $\{x : x = \alpha \mathbf{1}_E, \alpha > 0\}$ at any point along that ray, then Lovász extension is surface plot with coordinates on that 2D plane, or alternatively we can view contours (which we will do).

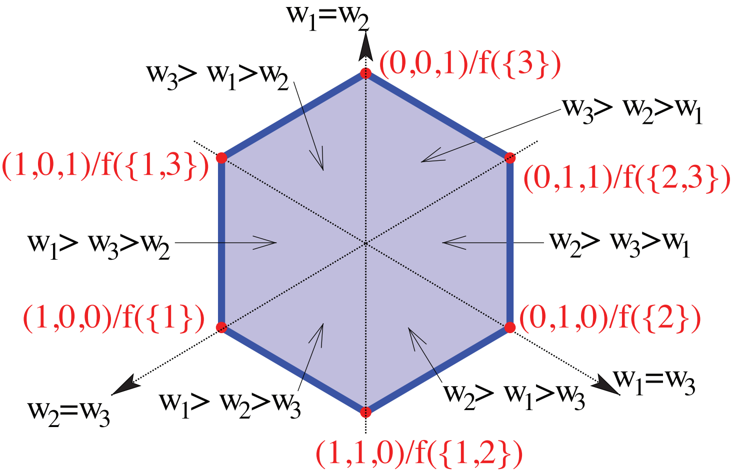
Example: $m = 3$, $E = \{1, 2, 3\}$

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



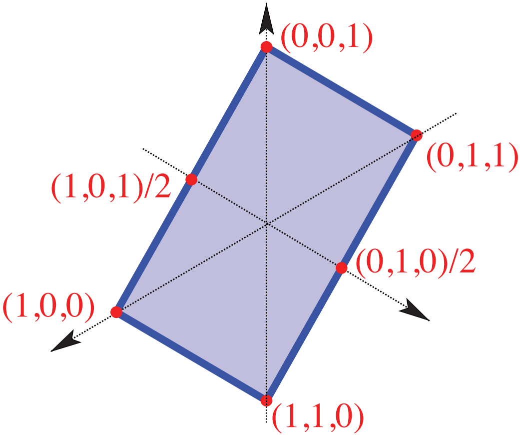
Example: $m = 3$, $E = \{1, 2, 3\}$

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



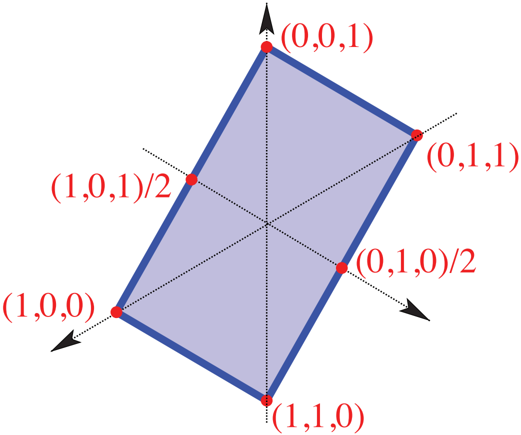
Example: $m = 3$, $E = \{1, 2, 3\}$

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



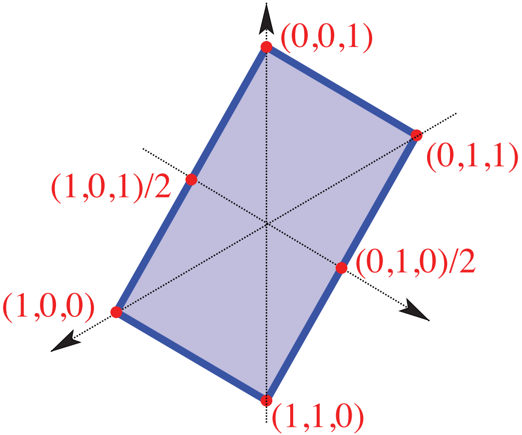
Example: $m = 3$, $E = \{1, 2, 3\}$

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



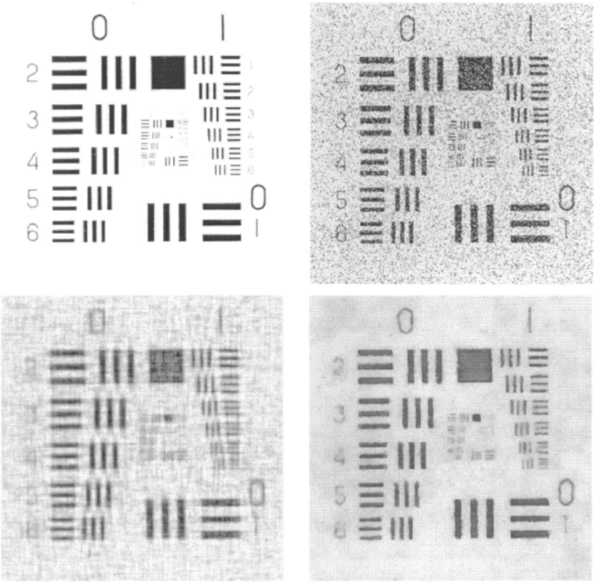
Example: $m = 3$, $E = \{1, 2, 3\}$

- Example 2 (from Bach-2011): $f(A) = |\mathbf{1}_{1\in A} - \mathbf{1}_{2\in A}| + |\mathbf{1}_{2\in A} - \mathbf{1}_{3\in A}|$
- This gives a "total variation" function for the Lovász extension, with $\tilde{f}(w) = |w_1 - w_2| + |w_2 - w_3|$.
- When used as a prior, prefers piecewise-constant signals (e.g., $\sum_i |w_i - w_{i+1}|$).



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 $M_j = \sum_{i=1}^{j} m(e_i)$, with $w(e_1) \geq w(e_2) \geq \ldots$ for some w.

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 $M_j = \sum_{i=1}^{j} m(e_i)$, with $w(e_1) \geq w(e_2) \geq \dots$ for some w.
- $\tilde{f}(w)$ is given as

$$\tilde{f}(w) = \sum_{i=1}^{m} w(e_i)\big(g(M_i) - g(M_{i-1})\big) \tag{17.58}$$

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 $M_j = \sum_{i=1}^{j} m(e_i)$, with $w(e_1) \geq w(e_2) \geq \dots$ for some w.
- $\tilde{f}(w)$ is given as

$$\tilde{f}(w) = \sum_{i=1}^{m} w(e_i)\big(g(M_i) - g(M_{i-1})\big) \tag{17.58}$$

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

$$\tilde{f}(w) = \sum_{i=1}^{m} w(e_i)\big(g(i) - g(i-1)\big) \tag{17.59}$$

Example: Lovász extension and cut functions

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

Example: Lovász extension and cut functions

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

$$f(X) = \sum_{i \in X, j \in V \setminus X} m_{ij} \tag{17.60}$$

Example: Lovász extension and cut functions

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

$$f(X) = \sum_{i \in X, j \in V \setminus X} m_{ij} \tag{17.60}$$

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

$$\tilde{f}(w) = \sum_{i,j \in V} m_{ij} \max \{(w_i - w_j), 0\} \tag{17.61}$$

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

Example: Lovász extension and cut functions

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

$$f(X) = \sum_{i \in X, j \in V \setminus X} m_{ij} \tag{17.60}$$

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

$$\tilde{f}(w) = \sum_{i,j \in V} m_{ij} \max \{(w_i - w_j), 0\} \tag{17.61}$$

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

A few more Lovász extension examples

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

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

(thanks to K. Narayanan).

Supervised And Unsupervised Machine Learning

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

ERM

$$\min_{w \in \mathbb{R}^n} \frac{1}{m} \sum_{i=1}^m \ell(y_i, w^\intercal x_i) + \lambda \Omega(w), \tag{17.62}$$

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

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

$$\min_{w^1,\dots,w^k \in \mathbb{R}^n} \sum_{j=1}^k \frac{1}{m} \sum_{i=1}^m \ell(y_i^k, (w^k)^\intercal x_i) + \lambda \Omega(w^k), \tag{17.63}$$

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

$$\min_{x_1,\dots,x_m} \min_{w^1,\dots,w^k \in \mathbb{R}^n} \sum_{j=1}^k \frac{1}{m} \sum_{i=1}^m \ell(y_i^k, (w^k)^\intercal x_i) + \lambda \Omega(w^k), \tag{17.64}$$

Norms, sparse norms, and computer vision

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

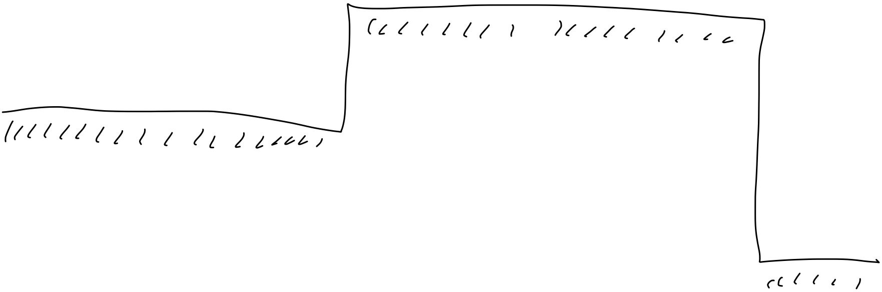
Norms, sparse norms, and computer vision

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

Norms, sparse norms, and computer vision

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

$$\Omega(w) = \sum_{i=2}^{N} |w_i - w_{i-1}| \tag{17.65}$$



Norms, sparse norms, and computer vision

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

$$\Omega(w) = \sum_{i=2}^{N} |w_i - w_{i-1}| \tag{17.65}$$

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

(Rodriguez, 2009)

Submodular parameterization of a sparse convex norm

- Prefer convex norms since they can be solved.

Submodular parameterization of a sparse convex norm

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

Submodular parameterization of a sparse convex norm

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

Submodular parameterization of a sparse convex norm

- Prefer convex norms since they can be solved.
- For $w \in \mathbb{R}^V$, $\text{supp}(w) \in \{0,1\}^V$ has $\text{supp}(w)(v) = 1$ iff $w(v) > 0$
- Desirable sparse norm: count the non-zeros, $\|w\|_0 = \mathbf{1}^\intercal \text{supp}(w)$.
- Using $\Omega(w) = \|w\|_0$ is NP-hard, instead we often optimize tightest convex relaxation, $\|w\|_1$ which is the convex envelope.

Submodular parameterization of a sparse convex norm

- Prefer convex norms since they can be solved.
- For $w \in \mathbb{R}^V$, $\text{supp}(w) \in \{0,1\}^V$ has $\text{supp}(w)(v) = 1$ iff $w(v) > 0$
- Desirable sparse norm: count the non-zeros, $\|w\|_0 = \mathbf{1}^\intercal \text{supp}(w)$.
- Using $\Omega(w) = \|w\|_0$ is NP-hard, instead we often optimize tightest convex relaxation, $\|w\|_1$ which is the convex envelope.
- With $\|w\|_0$ or its relaxation, each non-zero element has equal degree of penalty. Penalties do not interact.

$w = (0\ 0\ 0\ 0\ 0)$ $\quad$ $w = (1\ 0\ 1\ 1\ 1)$

$w = (0\ 1\ 0\ 0\ 0)$ $\quad$ $w = (1\ 1\ 1\ 1\ 1)$

Submodular parameterization of a sparse convex norm

- Prefer convex norms since they can be solved.
- For $w \in \mathbb{R}^V$, $\mathrm{supp}(w) \in \{0,1\}^V$ has $\mathrm{supp}(w)(v) = 1$ iff $w(v) > 0$
- Desirable sparse norm: count the non-zeros, $\|w\|_0 = \mathbf{1}^\intercal \mathrm{supp}(w)$.
- Using $\Omega(w) = \|w\|_0$ is NP-hard, instead we often optimize tightest convex relaxation, $\|w\|_1$ which is the convex envelope.
- With $\|w\|_0$ or its relaxation, each non-zero element has equal degree of penalty. Penalties do not interact.
- Given submodular function $f : 2^V \to \mathbb{R}_+$, $f(\mathrm{supp}(w))$ 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.

$\sqrt{\mathrm{supp}(w)}$

$\sqrt{\mathrm{supp}(w)(V_1)} + \sqrt{\mathrm{supp}(w)(V_2)}$

$f(A) = \sqrt{|A \cap V_1|} + \sqrt{|A \cap V_2|}$

Submodular parameterization of a sparse convex norm

- Prefer convex norms since they can be solved.
- For $w \in \mathbb{R}^V$, $\text{supp}(w) \in \{0,1\}^V$ has $\text{supp}(w)(v) = 1$ iff $w(v) > 0$
- Desirable sparse norm: count the non-zeros, $\|w\|_0 = \mathbf{1}^\intercal \text{supp}(w)$.
- Using $\Omega(w) = \|w\|_0$ is NP-hard, instead we often optimize tightest convex relaxation, $\|w\|_1$ which is the convex envelope.
- With $\|w\|_0$ or its relaxation, each non-zero element has equal degree of penalty. Penalties do not interact.
- Given submodular function $f : 2^V \to \mathbb{R}_+$, $f(\text{supp}(w))$ measures the "complexity" of the non-zero pattern of w; can have more non-zero values if they cooperate (via f) with other non-zero values.
- $f(\text{supp}(w))$ is hard to optimize, but it's convex envelope $\tilde{f}(|w|)$ (i.e., largest convex under-estimator of $f(\text{supp}(w))$) is obtained via the Lovász-extension $\tilde{f}$ of f (Vondrák 2007, Bach 2010).

Submodular parameterization of a sparse convex norm

- Prefer convex norms since they can be solved.
- For $w \in \mathbb{R}^V$, $\mathrm{supp}(w) \in \{0,1\}^V$ has $\mathrm{supp}(w)(v) = 1$ iff $w(v) > 0$
- Desirable sparse norm: count the non-zeros, $\|w\|_0 = \mathbf{1}^\intercal \mathrm{supp}(w)$.
- Using $\Omega(w) = \|w\|_0$ is NP-hard, instead we often optimize tightest convex relaxation, $\|w\|_1$ which is the convex envelope.
- With $\|w\|_0$ or its relaxation, each non-zero element has equal degree of penalty. Penalties do not interact.
- Given submodular function $f : 2^V \to \mathbb{R}_+$, $f(\mathrm{supp}(w))$ 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.
- $f(\mathrm{supp}(w))$ is hard to optimize, but it's convex envelope $\tilde{f}(|w|)$ (i.e., largest convex under-estimator of $f(\mathrm{supp}(w))$) is obtained via the Lovász-extension $\tilde{f}$ of f (Vondrák 2007, Bach 2010).
- Submodular functions thus parameterize structured convex sparse norms via the Lovász-extension!

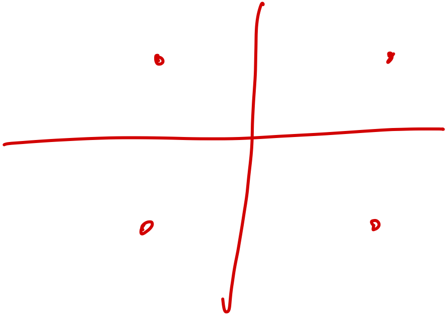
Submodular parameterization of a sparse convex norm

- Prefer convex norms since they can be solved.
- For $w \in \mathbb{R}^V$, $\text{supp}(w) \in \{0,1\}^V$ has $\text{supp}(w)(v) = 1$ iff $w(v) > 0$
- Desirable sparse norm: count the non-zeros, $\|w\|_0 = \mathbf{1}^\intercal \text{supp}(w)$.
- Using $\Omega(w) = \|w\|_0$ is NP-hard, instead we often optimize tightest convex relaxation, $\|w\|_1$ which is the convex envelope.
- With $\|w\|_0$ or its relaxation, each non-zero element has equal degree of penalty. Penalties do not interact.
- Given submodular function $f : 2^V \to \mathbb{R}_+$, $f(\text{supp}(w))$ measures the "complexity" of the non-zero pattern of w; can have more non-zero values if they cooperate (via f) with other non-zero values.
- $f(\text{supp}(w))$ is hard to optimize, but it's convex envelope $\tilde{f}(|w|)$ (i.e., largest convex under-estimator of $f(\text{supp}(w))$) is obtained via the Lovász-extension $\tilde{f}$ of f (Vondrák 2007, Bach 2010).
- Submodular functions thus parameterize structured convex sparse norms via the Lovász-extension!
- Ex: total variation is Lovász-ext. of graph cut, but $\exists$ many more!

Lovász extension and norms

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

• Note, $\|w\|_{\tilde{f}}$ is an actual norm ?



Lovász extension and norms

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

Lovász extension and norms

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

Lovász extension and norms

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