Submodular Functions, Optimization, and Applications to Machine Learning

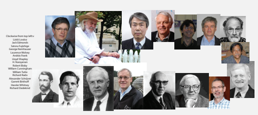
— Spring Quarter, Lecture 15 —

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

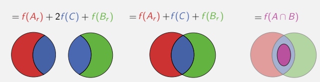
Prof. Jeff Bilmes

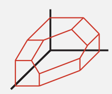
University of Washington, Seattle
Department of Electrical Engineering
http://melodi.ee.washington.edu/~bilmes

May 16th, 2018



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





Cumulative Outstanding Reading

- Read chapter 1 from Fujishige's book.
- Read chapter 2 from Fujishige's book.
- Read chapter 3 from Fujishige's book.
- Read chapter 4 from Fujishige's book.

Announcements, Assignments, and Reminders

- Next homework will be posted soon.
- As always, if you have any questions about anything, please ask then via our discussion board (https://canvas.uw.edu/courses/1216339/discussion_topics). Can meet at odd hours via zoom (send message on canvas to schedule time to chat).

Class Road Map - EE563

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

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

Continuous Extensions of Discrete Set Functions

- Any function $f : 2^V \to \mathbb{R}$ (equivalently $f : \{0,1\}^V \to \mathbb{R}$) can be extended to a continuous function in the sense $\tilde{f} : [0,1]^V \to \mathbb{R}$.
- This may be tight (i.e., $\tilde{f}(\mathbf{1}_A) = f(A)$ for all A). I.e., the extension $\tilde{f}$ coincides with f at the hypercube vertices.
- In fact, any such discrete function defined on the vertices of the n-D hypercube $\{0,1\}^n$ has a variety of both convex and concave extensions tight at the vertices (Crama & Hammer'11). Example $n = 1$,

Concave Extensions $\tilde{f} : [0,1] \to \mathbb{R}$ | Discrete Function $f : \{0,1\}^V \to \mathbb{R}$ | Convex Extensions $\tilde{f} : [0,1] \to \mathbb{R}$

- Since there are an exponential number of vertices $\{0,1\}^n$, important questions regarding such extensions is:
  1. When are they computationally feasible to obtain or estimate?
  2. When do they have nice mathematical properties?
  3. When are they useful for something practical?

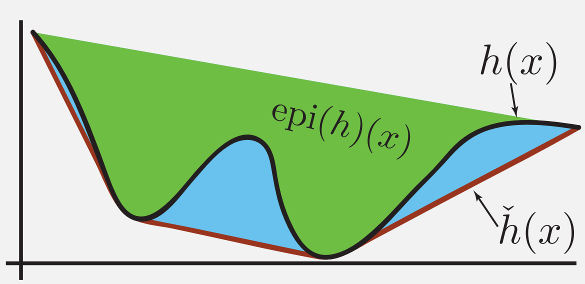
Def: Convex Envelope of a function

- Given any function $h : \mathbb{R}^n \to \mathbb{R}$, define new function $\breve{h} : \mathbf{R}^n \to \mathbb{R}$ via:

$$\breve{h}(x) = \sup \{g(x) : g \text{ is convex } \& \ g(y) \le h(y), \forall y \in \mathbb{R}^n\} \quad (15.6)$$

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

$$\breve{h}(x) = \inf \{t : (x,t) \in \text{convexhull}(\text{epigraph}(h))\} \quad (15.7)$$



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

$$\check{f}(x) = \min_{p \in \triangle^n(x)} \sum_{S \subseteq V} p_S f(S) \quad (15.1)$$

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

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

Cont. Extensions | Lovász extension | Choquet Integration | Lovász extn., defs/props | Lovász extension examples

Convex Closure of Discrete Set Functions

- Given, $\check{f}(x) = \min_{p \in \triangle^n(x)} E_{S \sim p}[f(S)]$, there are several things we'd like to 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.

Tightness of Convex Closure

Lemma 15.3.1

$\forall A \subseteq V$, *we have* $\check{f}(\mathbf{1}_A) = f(A)$.

Proof.

- Define p^x to be an achiving argmin in $\check{f}(x) = \min_{p \in \triangle^n(x)} E_{S \sim p}[f(S)]$.
- Take an arbitrary A, so that $\mathbf{1}_A = \sum_{S \subseteq V} p_S^{\mathbf{1}_A} \mathbf{1}_S = \mathbf{1}_A$.
- Suppose $\exists S'$ with $S' \setminus A \neq 0$ having $p_{S'}^{\mathbf{1}_A} > 0$. This would mean, for any $v \in S' \setminus A$, that $\left(\sum_S p_S^{\mathbf{1}_A} \mathbf{1}_S\right)(v) > 0$, a contradiction.
- Suppose $\exists S'$ s.t. $A \setminus S' \neq \emptyset$ with $p_{S'}^{\mathbf{1}_A} > 0$.
- Then, for any $v \in A \setminus S'$, consider below leading to a contradiction

$$\underbrace{p_{S'} \mathbf{1}_{S'}}_{>0} + \underbrace{\sum_{\substack{S \subseteq A \\ S \neq S'}} p_S \mathbf{1}_S}_{\text{can't sum to 1}} \Rightarrow \Big(\sum_{\substack{S \subseteq A \\ S \neq S'}} p_s \mathbf{1}_S\Big)(v) < 1 \tag{15.2}$$

I.e., $v \in A$ so it must get value 1, but since $v \notin S'$, v is deficient.

Convexity of the Convex Closure

Lemma 15.3.2

$\check{f}(x) = \min_{p \in \triangle^n(x)} E_{S \sim p}[f(S)]$ *is convex in* $[0,1]^V$.

Proof.

- Let $x, y \in [0,1]^V$, $0 \leq \lambda \leq 1$, and $z = \lambda x + (1-\lambda) y$, then

$$\lambda \check{f}(x) + (1-\lambda)\check{f}(y) = \lambda \sum_S p_S^x f(S) + (1-\lambda) \sum_S p_S^y f(S) \tag{15.3}$$

$$= \sum_S (\lambda p_S^x + (1-\lambda) p_S^y) f(S) \tag{15.4}$$

$$= \sum_S p_S^{z'} f(S) \geq \min_{p \in \triangle^n(z)} E_{S \sim p}[f(S)] \tag{15.5}$$

$$= \check{f}(z) = \check{f}(\lambda x + (1-\lambda) y) \tag{15.6}$$

- Note that $p_S^{z'} = \lambda p_S^x + (1-\lambda) p_S^y$ and is feasible in the min since $\sum_S p_S^{z'} = 1$, $p_S^{z'} \geq 0$ and $\sum_S p_S^{z'} \mathbf{1}_S = z$.

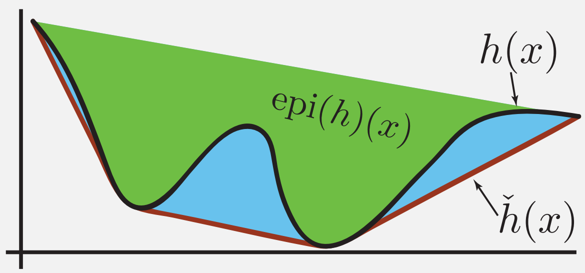
Def: Convex Envelope of a function

- Given any function $h : \mathbb{R}^n \to \mathbb{R}$, define new function $\check{h} : \mathbf{R}^n \to \mathbb{R}$ via:

$$\check{h}(x) = \sup \{g(x) : g \text{ is convex \& } g(y) \leq h(y), \forall y \in \mathbb{R}^n\} \qquad (15.6)$$

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

$$\check{h}(x) = \inf \{t : (x,t) \in \text{convexhull}(\text{epigraph}(h))\} \qquad (15.7)$$



Convex Closure is the Convex Envelope

Lemma 15.3.3

$\check{f}(x) = \min_{p \in \triangle^n(x)} E_{S\sim p}[f(S)]$ *is the convex envelope.*

Proof.

- Suppose $\exists$ a convex $\bar{f}$ with $\bar{f}(\mathbf{1}_A) = f(A) = \check{f}(\mathbf{1}_A), \forall A \subseteq V$ and $\exists x \in [0,1]^V$ s.t. $\bar{f}(x) > \check{f}(x)$.
- Define p^x to be an achiving argmin in $\check{f}(x) = \min_{p \in \triangle^n(x)} E_{S\sim p}[f(S)]$. Hence, we have $x = \sum_S p^x_S \mathbf{1}_S$. Thus

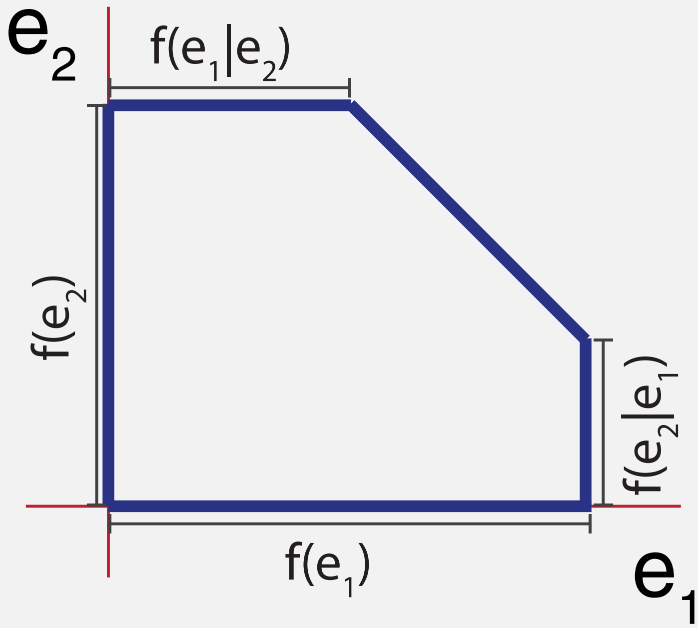
$$\check{f}(x) = \sum_S p^x_S f(S) = \sum_S p^x_S \bar{f}(\mathbf{1}_S) \qquad (15.7)$$

$$< \bar{f}(x) = \bar{f}(\sum_S p^x_S \mathbf{1}_S) \qquad (15.8)$$

but this contradicts the convexity of $\bar{f}$.

Polymatroid with labeled edge lengths

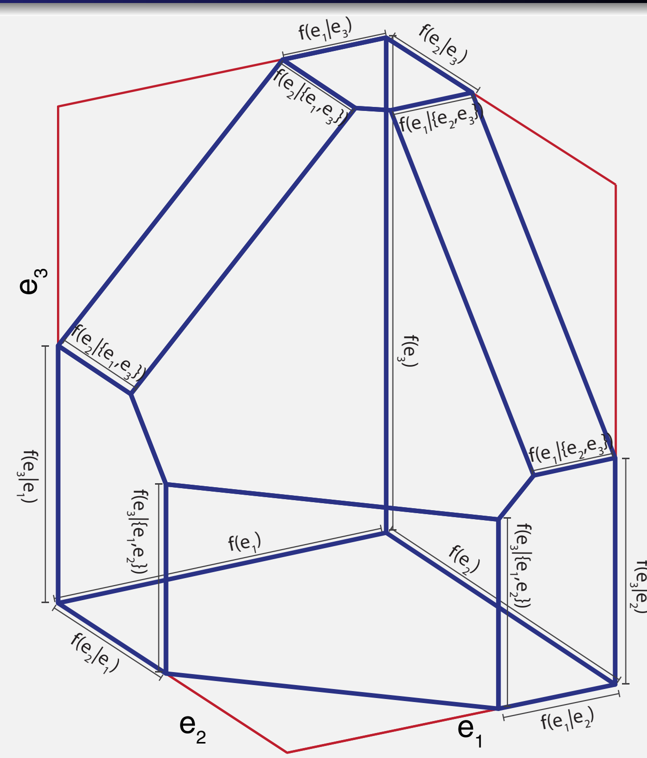
- Recall $f(e|A) = f(A+e) - f(A)$
- Notice how submodularity, $f(e|B) \leq f(e|A)$ for $A \subseteq B$, defines the shape of the polytope.
- In fact, we have strictness here $f(e|B) < f(e|A)$ for $A \subset B$.
- Also, consider how the greedy algorithm proceeds along the edges of the polytope.



Cont. Extensions | Lovász extension | Choquet Integration | Lovász extn., defs/props | Lovász extension examples

Polymatroid with labeled edge lengths

- Recall $f(e|A) = f(A+e) - f(A)$
- Notice how submodularity, $f(e|B) \leq f(e|A)$ for $A \subseteq B$, defines the shape of the polytope.
- In fact, we have strictness here $f(e|B) < f(e|A)$ for $A \subset B$.
- Also, consider how the greedy algorithm proceeds along the edges of the polytope.



Optimization over P_f

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

$$\begin{aligned} &\text{maximize} && w^\intercal x && (15.9a)\\ &\text{subject to} && x \in P_f && (15.9b) \end{aligned}$$

- Since P_f is down closed, if $\exists e \in E$ with $w(e) < 0$ then the solution above is unboundedly large. Hence, assume $w \in \mathbb{R}_+^E$.
- Due to Theorem **??**, any $x \in P_f$ with $x \notin B_f$ is dominated by $x \leq y \in B_f$ which can only increase $w^\intercal x \leq w^\intercal y$ when $w \in \mathbb{R}_+^E$.
- Hence, the problem is equivalent to: given $w \in \mathbb{R}_+^E$,

$$\begin{aligned} &\text{maximize} && w^\intercal x && (15.10a)\\ &\text{subject to} && x \in B_f && (15.10b) \end{aligned}$$

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

A continuous extension of f

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

$$\begin{aligned} &\text{maximize} && w^\intercal x && (15.11a)\\ &\text{subject to} && x \in B_f && (15.11b) \end{aligned}$$

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

$$\breve{f}(w) = \max(wx : x \in B_f) \quad (15.12)$$

- Hence, for any w, from the solution to the above theorem (as we have seen), we can compute the value of this function using Edmond's greedy algorithm.

Edmond's Theorem: The Greedy Algorithm

- Edmonds proved that the solution to $\breve{f}(w) = \max(wx : x \in B_f)$ is solved by the greedy algorithm iff f is submodular.
- In particular, sort choose element order $(e_1, e_2, \ldots, e_m)$ based on decreasing w,so that $w(e_1) \geq w(e_2) \geq \cdots \geq w(e_m)$.
- Define the chain with i^{th} element $E_i = \{e_1, e_2, \ldots, e_i\}$.
- Define a vector $x^* \in \mathbb{R}^V$ where element e_i has value $x(e_i) = f(e_i|E_{i-1})$ for all $i \in V$.
- Then $\langle w, x^* \rangle = \max(wx : x \in B_f)$

Theorem 15.4.1 (Edmonds)

If $f : 2^E \to \mathbb{R}_+$ is given, and B is a polytope in $\mathbb{R}^E_+$ of the form $B = \left\{x \in \mathbb{R}^E_+ : x(A) \leq f(A), \forall A \subseteq E, x(E) = f(E)\right\}$, then the greedy solution to the problem $\max(w^\intercal x : x \in P)$ is $\forall w$ optimum iff f is monotone non-decreasing submodular (i.e., iff P is a polymatroid).

A continuous extension of submodular f

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

$$\breve{f}(w) = \max(wx : x \in B_f) \tag{15.13}$$

$$= \sum_{i=1}^{m} w(e_i) f(e_i|E_{i-1}) = \sum_{i=1}^{m} w(e_i) x(e_i) \tag{15.14}$$

$$= \sum_{i=1}^{m} w(e_i)(f(E_i) - f(E_{i-1})) \tag{15.15}$$

$$= w(e_m) f(E_m) + \sum_{i=1}^{m-1} (w(e_i) - w(e_{i+1})) f(E_i) \tag{15.16}$$

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

A continuous extension of submodular f

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

$$\breve{f}(w) = \max(wx : x \in B_f) \tag{15.17}$$

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

$$\breve{f}(w) = w(e_m)f(E_m) + \sum_{i=1}^{m-1}(w(e_i) - w(e_{i+1}))f(E_i) \tag{15.18}$$

$$= \sum_{i=1}^{m} \lambda_i f(E_i) \tag{15.19}$$

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

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

An extension of f

- Recall, for any such $w \in \mathbb{R}^E$, we have

$$\begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} = \underbrace{(w_1 - w_2)}_{\lambda_1} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \underbrace{(w_2 - w_3)}_{\lambda_2} \begin{pmatrix} 1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \cdots + \underbrace{(w_{n-1} - w_n)}_{\lambda_{m-1}} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 0 \end{pmatrix} + \underbrace{(w_m)}_{\lambda_m} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{pmatrix} \tag{15.20}$$

- If we take w in decreasing order, then each coefficient of the vectors is non-negative (except possibly the last one, $\lambda_m = w_m$).
- Often, we take $w \in \mathbb{R}_+^V$ or even $w \in [0,1]^V$, where $\lambda_m \geq 0$.

An extension of f

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

$$E_i \stackrel{\text{def}}{=} \{e_1, e_2, \dots, e_i\} \tag{15.21}$$

- Note that

$$\mathbf{1}_{E_0} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \mathbf{1}_{E_1} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \dots, \mathbf{1}_{E_\ell} = \begin{pmatrix} \left.\begin{matrix} 1 \\ 1 \\ \vdots \\ 1 \end{matrix}\right\} \ell\times \\ \left.\begin{matrix} 0 \\ 0 \\ \vdots \\ 0 \end{matrix}\right\} (n-\ell)\times \end{pmatrix}, \text{ etc.}$$

- Hence, from the previous and current slide, we have $w = \sum_{i=1}^m \lambda_i \mathbf{1}_{E_i}$

From $\breve{f}$ back to f, even when f is not submodular

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

$$w = (w(e_1), w(e_2), \dots, w(e_m)) = (\underbrace{1, 1, 1, \dots, 1}_{|A| \text{ times}}, \underbrace{0, 0, \dots, 0}_{m-|A| \text{ times}}) \tag{15.22}$$

so that $1_A(i) = 1$ if $i \leq |A|$, and $1_A(i) = 0$ otherwise.

- For any $f : 2^E \to \mathbb{R}$, $w = \mathbf{1}_A$, since $E_{|A|} = \{e_1, e_2, \dots, e_{|A|}\} = A$:

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

$$= \mathbf{1}_A(m) f(E_m) + \sum_{i=1}^{m-1} (\mathbf{1}_A(i) - \mathbf{1}_A(i+1)) f(E_i) \tag{15.23}$$

$$= (\mathbf{1}_A(|A|) - \mathbf{1}_A(|A|+1)) f(E_{|A|}) = f(E_{|A|}) = f(A) \tag{15.24}$$

From $\breve{f}$ back to f

- We can view $\breve{f} : [0,1]^E \to \mathbb{R}$ defined on the hypercube, with f defined as $\breve{f}$ evaluated on the hypercube extreme points (vertices).
- To summarize, with $\breve{f}(\mathbf{1}_A) = \sum_{i=1}^m \lambda_i f(E_i)$, we have

$$\breve{f}(\mathbf{1}_A) = f(A), \tag{15.25}$$

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

$$\breve{f}(\mathbf{1}_A) = \max \{\mathbf{1}_A{}^\intercal x : x \in B_f\} \tag{15.26}$$

$$= \max \{\mathbf{1}_A{}^\intercal x : x(B) \leq f(B), \forall B \subseteq E\} \tag{15.27}$$

- Note when considering only $\breve{f} : [0,1]^E \to \mathbb{R}$, then any $w \in [0,1]^E$ is in positive orthant, and we have

$$\breve{f}(w) = \max \{w^\intercal x : x \in P_f\} \tag{15.28}$$

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

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

$$\breve{f}(w) = \sum_{i=1}^m \lambda_i f(E_i) \tag{15.29}$$

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

$$\text{for } i \in \{1, \ldots, m\}, \quad \lambda_i = \begin{cases} w(e_i) - w(e_{i+1}) & \text{if } i < m \\ w(e_m) & \text{if } i = m \end{cases} \tag{15.30}$$

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!

Weighted gains vs. weighted functions

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

$$\breve{f}(w) = \sum_{i=1}^{m} w(e_i) f(e_i | E_{i-1}) \quad (15.31)$$

$$= \sum_{i=1}^{m} w(e_i)(f(E_i) - f(E_{i-1})) \quad (15.32)$$

$$= w(e_m) f(E_m) + \sum_{i=1}^{m-1} (w(e_i) - w(e_{i+1})) f(E_i) \quad (15.33)$$

$$= \sum_{i=1}^{m} \lambda_i f(E_i) \quad (15.34)$$

- So $\breve{f}(w)$ seen either as sum of weighted gain evaluations (Eqn. (15.31)), or as sum of weighted function evaluations (Eqn. (15.34)).

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:

$$\breve{f}(w) = \max(wx : x \in B_f) = \sum_{i=1}^{m} \lambda_i f(E_i) \quad (15.35)$$

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

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

$$\breve{f}(w) = \sum_{i=1}^{m} \lambda_i f(E_i) \quad (15.36)$$

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

- In both Eq. (15.35) and Eq. (15.36), we have $\breve{f}(\mathbf{1}_A) = f(A)$, $\forall A$, but Eq. (15.36), might not be convex.
- Submodularity is sufficient for convexity, but is it necessary?

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

- Lovász showed that if a function $\breve{f}(w)$ defined as in Eqn. (15.29) is convex, then f must be submodular.
- This continuous extension $\breve{f}$ of f, in any case (f being submodular or not), is typically called the Lovász extension of f (but also sometimes called the Choquet integral, or the Lovász-Edmonds extension).

Lovász Extension, Submodularity and Convexity

Theorem 15.4.2

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.(15.29) 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 Extension, Submodularity and Convexity

. . . proof of Thm. 15.4.2 cont.

- Earlier, we saw that $\breve{f}(\mathbf{1}_A) = f(A)$ for all $A \subseteq E$.
- Now, given $A, B \subseteq E$, we will show that

$$\breve{f}(\mathbf{1}_A + \mathbf{1}_B) = \breve{f}(\mathbf{1}_{A\cup B} + \mathbf{1}_{A\cap B}) \tag{15.37}$$
$$= f(A \cup B) + f(A \cap B). \tag{15.38}$$

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

$$w = (w(e_1), w(e_2), \dots, w(e_m)) \tag{15.39}$$
$$= (\underbrace{2, 2, \dots, 2}_{i\in C}, \underbrace{1, 1, \dots, 1}_{i\in A\triangle B}, \underbrace{0, 0, \dots, 0}_{i\in E\setminus(A\cup B)}) \tag{15.40}$$

- Then, considering $\breve{f}(w) = \sum_i \lambda_i f(E_i)$, we have $\lambda_{|C|} = 1$, $\lambda_{|A\cup B|} = 1$, and $\lambda_i = 0$ for $i \notin \{|C|, |A \cup B|\}$.
- But then $E_{|C|} = A \cap B$ and $E_{|A\cup B|} = A \cup B$. Therefore, $\breve{f}(w) = \breve{f}(\mathbf{1}_A + \mathbf{1}_B) = f(A \cap B) + f(A \cup B)$.

. . .

Lovász Extension, Submodularity and Convexity

. . . proof of Thm. 15.4.2 cont.

- Also, since $\breve{f}$ is convex (by assumption) and positively homogeneous, we have for any $A, B \subseteq E$,

$$0.5[f(A \cap B) + f(A \cup B)] = 0.5[\breve{f}(\mathbf{1}_A + \mathbf{1}_B)] \tag{15.41}$$
$$= \breve{f}(0.5\mathbf{1}_A + 0.5\mathbf{1}_B) \tag{15.42}$$
$$\leq 0.5\breve{f}(\mathbf{1}_A) + 0.5\breve{f}(\mathbf{1}_B) \tag{15.43}$$
$$= 0.5(f(A) + f(B)) \tag{15.44}$$

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

$$f(A \cup B) + f(A \cap B) \leq f(A) + f(B) \tag{15.45}$$

so f must be submodular.

□

Lovász ext. vs. the concave closure of submodular function

- The above theorem showed that the Lovász extension is convex iff f is submodular.
- Our next theorem shows that the Lovász extension coincides precisely with the convex closure iff f is submodular.
- I.e., not only is the Lovász extension convex for f submodular, it is the convex closure when f is convex.
- Hence, convex closure is easy to evaluate when f is submodular and is this particular form iff f is submodular.

Lovász ext. vs. the concave closure of submodular function

Theorem 15.4.3

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

Proof.

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

$$\bar{p}^x_A \leftarrow p^x_A - p^x_B \qquad \bar{p}^x_B \leftarrow p^x_B - p^x_B \tag{15.46}$$

$$\bar{p}^x_{A\cup B} \leftarrow p^x_{A\cup B} + p^x_B \qquad \bar{p}^x_{A\cap B} \leftarrow p^x_{A\cap B} + p^x_B \tag{15.47}$$

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

...

Lovász ext. vs. the concave closure of submodular function

. . . proof cont.

- This does increase $\sum_S p^x_S |S|^2$ however since

$$|A \cup B|^2 + |A \cap B|^2 = (|A| + |B \setminus A|)^2 + (|B| - |B \setminus A|)^2 \quad (15.48)$$
$$= |A|^2 + |B|^2 + 2|B \setminus A|(|A| - |B| + |B \setminus A|) \quad (15.49)$$
$$\geq |A|^2 + |B|^2 \quad (15.50)$$

- Contradiction! Hence, there can be no crossing sets A, B and we must have, for any A, B with $p^x_A > 0$ and $p^x_B > 0$ either $A \subset B$ or $B \subset A$.
- Hence, the sets $\{A \subseteq V : p^x_A > 0\}$ form a chain and can be as large only as size $n = |V|$.
- This is the same chain that defines the Lovász extension $\breve{f}(x)$, namely $\emptyset = E_0 \subseteq E_1 \subseteq E_2 \subset \ldots$ where $E_i = \{e_1, e_2, \ldots, e_i\}$ and e_i is orderd so that $x(e_1) \geq x(e_2) \geq \cdots \geq x(e_n)$.

Lovász ext. vs. the concave closure of submodular function

. . . proof cont.

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

$$\breve{f}(x) \leq \frac{1}{2}[f(S + i) + f(S + j)] < \breve{f}(x) \quad (15.51)$$

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

Integration and Aggregation

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

$$\int_X f du = \sup I_X(s) \tag{15.52}$$

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

Integration, Aggregation, and Weighted Averages

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

$$\text{WAVG}(x) = \sum_{e \in E} x(e)w(e) \tag{15.53}$$

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

$$\text{WAVG}(\mathbf{1}_e) = w(e) \tag{15.54}$$

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

$$\text{WAVG}(x) = \sum_{e \in E} x(e)w(e) = \sum_{e \in E} x(e)\text{WAVG}(\mathbf{1}_e) \tag{15.55}$$

Integration, Aggregation, and Weighted Averages

$$\text{WAVG}(x) = \sum_{e \in E} x(e)w(e) \tag{15.56}$$

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

$$\text{WAVG}_{w_1+w_2}(x) = \text{WAVG}_{w_1}(x) + \text{WAVG}_{w_2}(x), \tag{15.57}$$

$$\text{WAVG}_w(x_1 + x_2) = \text{WAVG}_w(x_1) + \text{WAVG}_w(x_2), \tag{15.58}$$

and,

$$\text{WAVG}(\alpha x) = \alpha \text{WAVG}(x). \tag{15.59}$$

- We will see: The Lovász extension is still be linear in "weights" (i.e., the submodular function f), but will not be linear in x and will only be positively homogeneous (for $\alpha \geq 0$).

Cont. Extensions | Lovász extension | Choquet Integration | Lovász extn., defs/props | Lovász extension examples

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 $A \subseteq E$):

$$\text{AG}(\mathbf{1}_A) = w_A \tag{15.60}$$

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

$$\text{AG}(x) = \sum_{A \subseteq E} x(A)w_A = \sum_{A \subseteq E} x(A)\text{AG}(\mathbf{1}_A) \tag{15.61}$$

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

$$\text{WAVG}_{w'}(x) = \text{AG}_w(x) \tag{15.62}$$

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

Integration, Aggregation, and Weighted Averages

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

Choquet integral

Definition 15.5.1

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

$$C_f(w) = \sum_{i=1}^{m} (w_{e_i} - w_{e_{i+1}}) f(E_i) \tag{15.63}$$

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:

$$C_f(w) = \sum_{i=1}^{m} w(e_i)(f(E_i) - f(E_{i-1})) \tag{15.64}$$

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

Choquet integral

Definition 15.5.1

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

$$C_f(w) = \sum_{i=1}^{m} (w_{e_i} - w_{e_{i+1}}) f(E_i) \qquad (15.63)$$

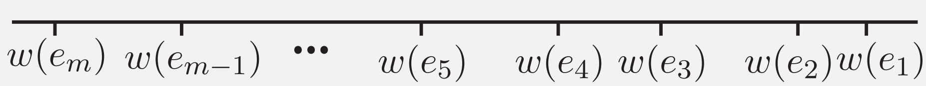
where in the sum, we have sorted and renamed the elements of E so that $w_{e_1} \geq w_{e_2} \geq \cdots \geq w_{e_m} \geq w_{e_{m+1}} \triangleq 0$, and where $E_i = \{e_1, e_2, \ldots, e_i\}$.

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

$$\sum_{k=m}^{n} a_k b_k = \sum_{k=m}^{n} A_k(b_k - b_{k+1}) + A_n b_{n+1} - A_{m-1} b_m \qquad (15.65)$$

The "integral" in the Choquet integral

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



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

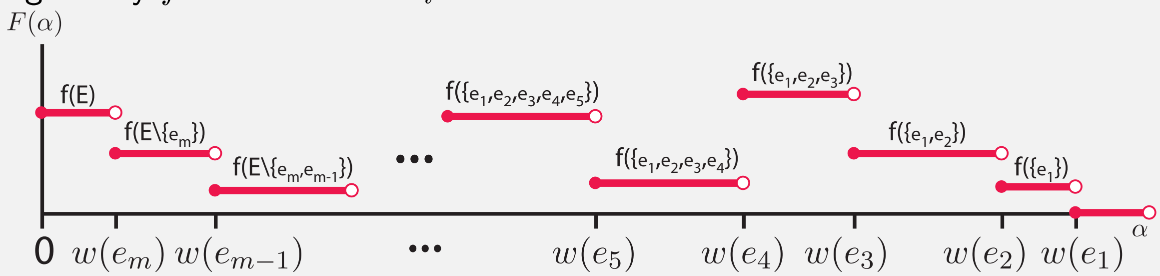
$$F_i(\alpha) = f(\{e \in E : w_e > \alpha\}) = f(E_i) \qquad (15.66)$$

The "integral" in the Choquet integral

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

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

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



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

The "integral" in the Choquet integral

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

$$\tilde{f}(w) \stackrel{\text{def}}{=} \int_0^\infty F(\alpha) d\alpha \quad (15.67)$$

$$= \int_0^\infty f(\{e \in E : w_e > \alpha\}) d\alpha \quad (15.68)$$

$$= \int_{w_{m+1}}^\infty f(\{e \in E : w_e > \alpha\}) d\alpha \quad (15.69)$$

$$= \sum_{i=1}^m \int_{w_{i+1}}^{w_i} f(\{e \in E : w_e > \alpha\}) d\alpha \quad (15.70)$$

$$= \sum_{i=1}^m \int_{w_{i+1}}^{w_i} f(E_i) d\alpha = \sum_{i=1}^m f(E_i)(w_i - w_{i+1}) \quad (15.71)$$

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 15.5.2

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

$$\tilde{f}(w) \stackrel{\text{def}}{=} \int_0^\infty F(\alpha) d\alpha \tag{15.72}$$

where the function F is defined as before.

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

Choquet integral and aggregation

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

$$\text{AG}(x) = \sum_{A \subseteq E} x(A) w_A = \sum_{A \subseteq E} x(A) \text{AG}(\mathbf{1}_A) \tag{15.73}$$

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

Choquet integral and aggregation

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

$$\alpha_0^x(A) + \sum_{j=1}^{m} \alpha_j^x(A) x_j \in \mathbb{R} \tag{15.74}$$

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

$$\mathsf{AG}(x) \stackrel{\text{def}}{=} \sum_{A:\mathbf{1}_A \in \mathcal{V}(x)} \Big(\alpha_0^x(A) + \sum_{j=1}^{m} \alpha_j^x(A) x_j\Big) \mathsf{AG}(\mathbf{1}_A) \tag{15.75}$$

Choquet integral and aggregation

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

$$\text{conv}(\mathcal{V}_\sigma) = \left\{x \in [0,1]^n | x_{\sigma(1)} \geq x_{\sigma(2)} \geq \cdots \geq x_{\sigma(m)}\right\} \tag{15.76}$$

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

Proposition 15.5.3

The above linear interpolation in Eqn. (15.75) *using the canonical partition yields the Lovász extension with* $\alpha_0^x(A) + \sum_{j=1}^m \alpha_j^x(A) x_j = x_{\sigma_i} - x_{\sigma_{i-1}}$ *for* $A = E_i = \{e_{\sigma_1}, \ldots, e_{\sigma_i}\}$ *for appropriate order* σ.

- Hence, Lovász extension is a generalized aggregation function.

Lovász extension as max over orders

- We can also write the Lovász extension as follows:

$$\tilde{f}(w) = \max_{\sigma \in \Pi_{[m]}} w^\intercal c^\sigma \quad (15.77)$$

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:

$$c_i^\sigma = f(E_{\sigma_i}) - f(E_{\sigma_{i-1}}) \quad (15.78)$$

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:

$$\tilde{f}(w) = \max_{x \in P_f} w^\intercal x = \max_{x \in B_f} w^\intercal x \quad (15.79)$$

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

Lovász extension, defined in multiple ways

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

$$\tilde{f}(w) = \sum_{i=1}^{m} w(e_i) f(e_i | E_{i-1}) \quad (15.80)$$

$$= \sum_{i=1}^{m-1} f(E_i)(w(e_i) - w(e_{i+1})) + f(E)w(e_m)a \quad (15.81)$$

$$= \sum_{i=1}^{m-1} \lambda_i f(E_i) \quad (15.82)$$

Lovász extension, as integral

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

$$\tilde{f}(w) = \sum_{i=1}^{m} w(e_i) f(e_i | E_{i-1}) = \sum_{i=1}^{m} \lambda_i f(E_i) \tag{15.83}$$

$$= \sum_{i=1}^{m-1} f(E_i)(w(e_i) - w(e_{i+1})) + f(E)w(e_m) \tag{15.84}$$

$$= \int_{\min\{w_1,\ldots,w_m\}}^{+\infty} f(\{w \geq \alpha\}) d\alpha + f(E) \min\{w_1, \ldots, w_m\} \tag{15.85}$$

$$= \int_{0}^{+\infty} f(\{w \geq \alpha\}) d\alpha + \int_{-\infty}^{0} [f(\{w \geq \alpha\}) - f(E)] d\alpha \tag{15.86}$$

general Lovász extension, as simple integral

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

$$\tilde{f}(w) = \int_{-\infty}^{+\infty} \hat{f}(\alpha) d\alpha \tag{15.87}$$

where

$$\hat{f}(\alpha) = \begin{cases} f(\{w \geq \alpha\}) & \text{if } \alpha >= 0 \\ f(\{w \geq \alpha\}) - f(E) & \text{if } \alpha < 0 \end{cases} \tag{15.88}$$

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

Lovász extension, as integral

- To show Eqn. (15.85), 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

$$f(\{w \geq \alpha\}) = \begin{cases} 0 & \text{if } \alpha > w(e_1) \\ f(E_k) & \text{if } \alpha \in (w(e_{k+1}), w(e_k)), k \in \{1, \ldots, m-1\} \\ f(E) & \text{if } \alpha < w(e_m) \end{cases} \tag{15.89}$$

  we may use open intervals since sets of zero measure don't change integration.
- Inside the integral, then, this recovers Eqn. (15.84).

Lovász extension, as integral

- To show Eqn. (15.86), start with Eqn. (15.85), 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. (15.86), i.e.:

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

Lovász extension properties

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

Theorem 15.6.1

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

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

Lovász extension properties: ex. property 3

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

Lovász extension properties

- Given Eqns. (15.83) through (15.86), most of the above properties are relatively easy to derive.
- For example, if f is symmetric, and since $f(E) = f(\emptyset) = 0$, we have

$$\tilde{f}(-w) = \int_{-\infty}^{\infty} f(\{-w \geq \alpha\})d\alpha = \int_{-\infty}^{\infty} f(\{w \leq -\alpha\})d\alpha \quad (15.90)$$

$$\stackrel{(a)}{=} \int_{-\infty}^{\infty} f(\{w \leq \alpha\})d\alpha \stackrel{(b)}{=} \int_{-\infty}^{\infty} f(\{w > \alpha\})d\alpha \quad (15.91)$$

$$= \int_{-\infty}^{\infty} f(\{w \geq \alpha\})d\alpha = \tilde{f}(w) \quad (15.92)$$

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

$$\tilde{f}(w) = \mathbb{E}_{p(\alpha)}[\underbrace{f(\{w \geq \alpha\})}_{h(\alpha)}] = \mathbb{E}_{p(\alpha)}[\underbrace{f(e \in E : w(e_i) \geq \alpha)}_{h(\alpha)}] \quad (15.93)$$

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

- Useful for showing results for randomized rounding schemes in solving submodular opt. problems subject to constraints via relaxations to convex optimization problems subject to linear constraints.

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

- If $w_1 \geq w_2$, then

$$\tilde{f}(w) = w_1 f(\{1\}) + w_2 f(\{2\}|\{1\}) \tag{15.94}$$
$$= (w_1 - w_2) f(\{1\}) + w_2 f(\{1, 2\}) \tag{15.95}$$

- If $w_1 \leq w_2$, then

$$\tilde{f}(w) = w_2 f(\{2\}) + w_1 f(\{1\}|\{2\}) \tag{15.96}$$
$$= (w_2 - w_1) f(\{2\}) + w_1 f(\{1, 2\}) \tag{15.97}$$

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

- If $w_1 \geq w_2$, then

$$\tilde{f}(w) = w_1 f(\{1\}) + w_2 f(\{2\}|\{1\}) \tag{15.98}$$
$$= (w_1 - w_2) f(\{1\}) + w_2 f(\{1, 2\}) \tag{15.99}$$
$$= \frac{1}{2} f(1)(w_1 - w_2) + \frac{1}{2} f(1)(w_1 - w_2) \tag{15.100}$$
$$+ \frac{1}{2} f(\{1, 2\})(w_1 + w_2) - \frac{1}{2} f(\{1, 2\})(w_1 - w_2) \tag{15.101}$$
$$+ \frac{1}{2} f(2)(w_1 - w_2) + \frac{1}{2} f(2)(w_2 - w_1) \tag{15.102}$$

- 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

$$\tilde{f}(w) = \frac{1}{2}\left(f(\{1\}) + f(\{2\}) - f(\{1,2\})\right)|w_1 - w_2| \tag{15.103}$$
$$+ \frac{1}{2}\left(f(\{1\}) - f(\{2\}) + f(\{1,2\})\right) w_1 \tag{15.104}$$
$$+ \frac{1}{2}\left(-f(\{1\}) + f(\{2\}) + f(\{1,2\})\right) w_2 \tag{15.105}$$
$$= -\left(f(\{1\}) + f(\{2\}) - f(\{1,2\})\right) \min\{w_1, w_2\} \tag{15.106}$$
$$+ f(\{1\})w_1 + f(\{2\})w_2 \tag{15.107}$$

- 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

$$\tilde{f}(w) = w_1 f(\{1\}) + w_2 f(\{2\}|\{1\}) \tag{15.108}$$

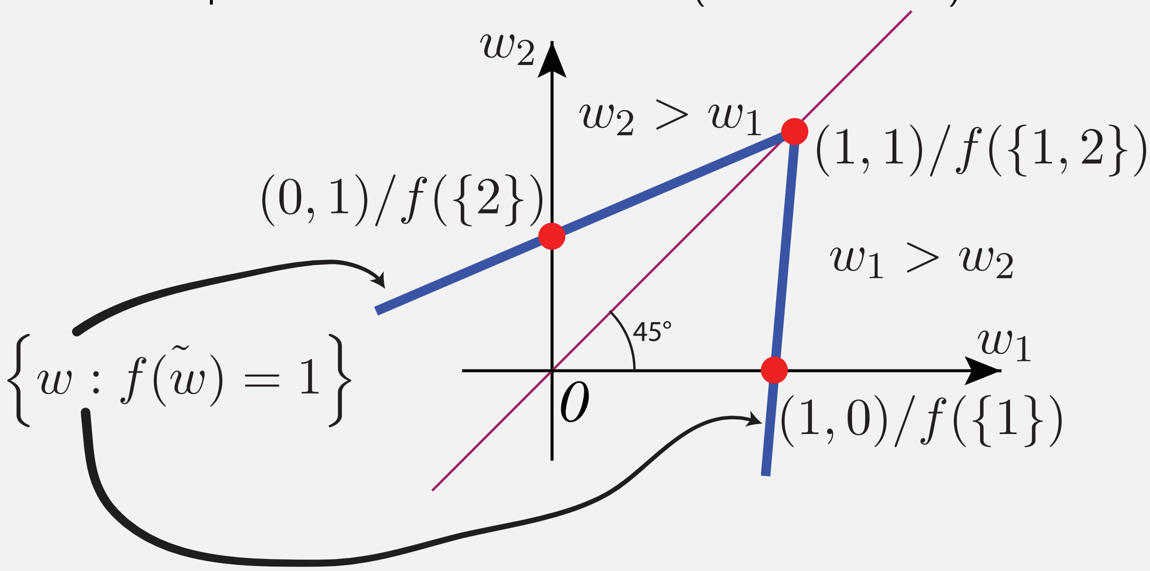
  - If $w = (1,0)/f(\{1\}) = \left(1/f(\{1\}), 0\right)$ then $\tilde{f}(w) = 1$.
  - If $w = (1,1)/f(\{1,2\})$ then $\tilde{f}(w) = 1$.
- If $w_1 \leq w_2$, then

$$\tilde{f}(w) = w_2 f(\{2\}) + w_1 f(\{1\}|\{2\}) \tag{15.109}$$

  - 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).



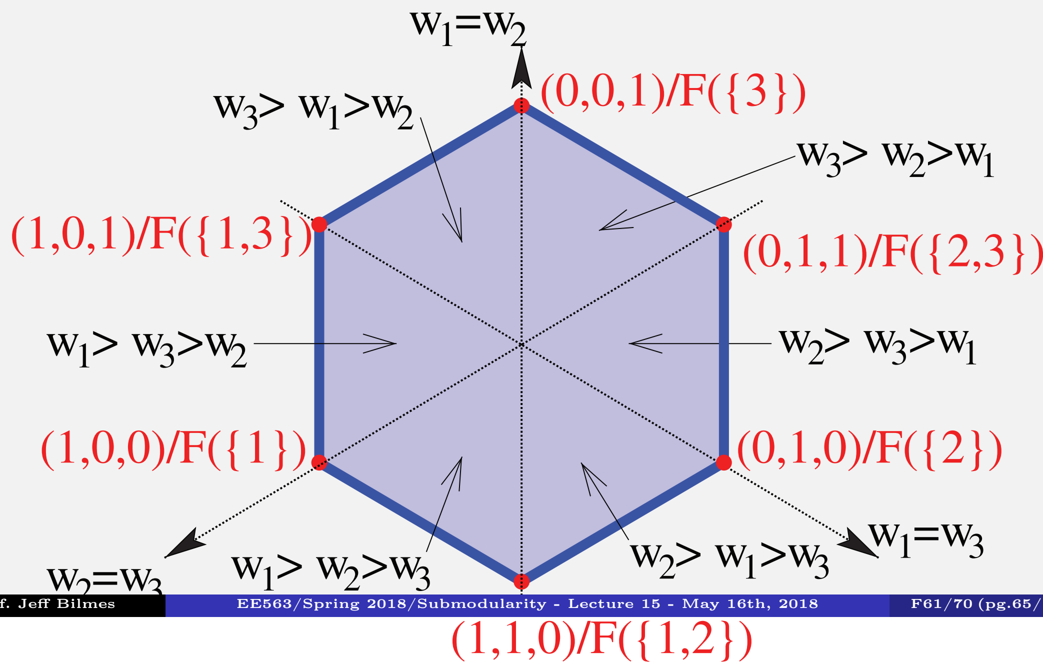
Cont. Extensions | Lovász extension | Choquet Integration | Lovász extn., defs/props | Lovász extension examples

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

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

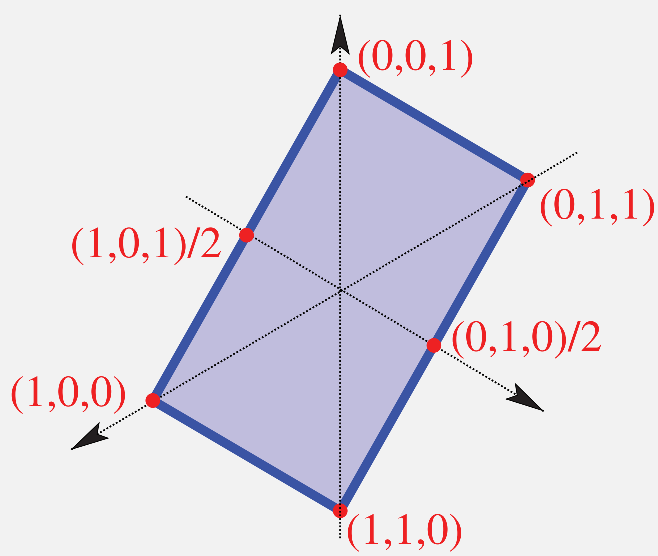
Example: $m = 3$, $E = \{1, 2, 3\}$

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



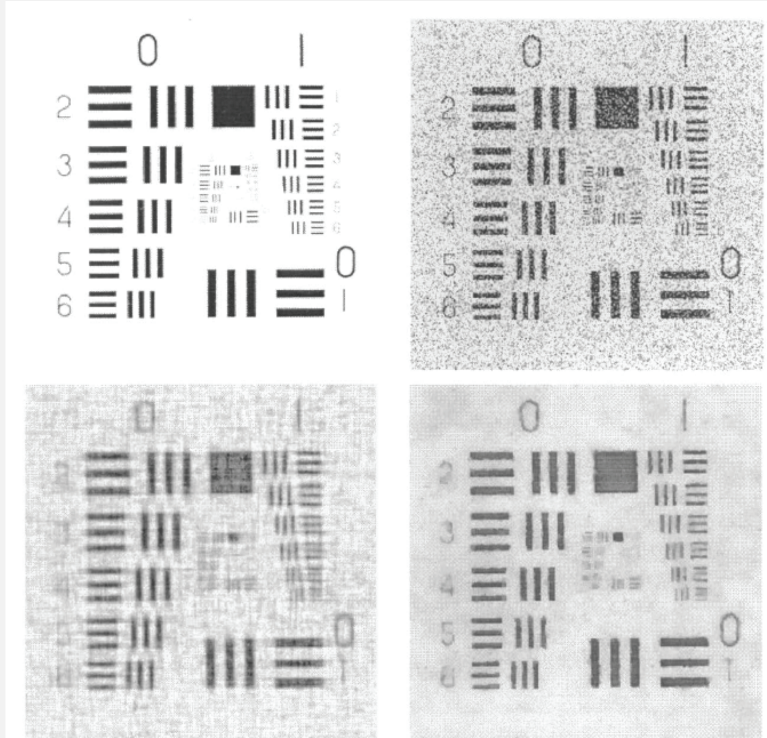
Example: $m = 3$, $E = \{1, 2, 3\}$

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



Total Variation Example

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



Example: Lovász extension of concave over modular

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

$$\tilde{f}(w) = \sum_{i=1}^{m} w(e_i)\big(g(M_i) - g(M_{i-1})\big) \tag{15.110}$$

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

$$\tilde{f}(w) = \sum_{i=1}^{m} w(e_i)\big(g(i) - g(i-1)\big) \tag{15.111}$$

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

$$f(X) = \sum_{i \in X, j \in V \setminus X} m_{ij} \tag{15.112}$$

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

$$\tilde{f}(w) = \sum_{i,j \in V} m_{ij} \max \{(w_i - w_j), 0\} \tag{15.113}$$

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

A few more Lovász extension examples

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

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

(thanks to K. Narayanan).

Supervised And Unsupervised Machine Learning

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

$$\min_{w \in \mathbb{R}^n} \frac{1}{m} \sum_{i=1}^{m} \ell(y_i, w^\intercal x_i) + \lambda \Omega(w), \tag{15.114}$$

  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:

$$\min_{w^1,\dots,w^k \in \mathbb{R}^n} \sum_{j=1}^{k} \frac{1}{m} \sum_{i=1}^{m} \ell(y_i^k, (w^k)^\intercal x_i) + \lambda \Omega(w^k), \tag{15.115}$$

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

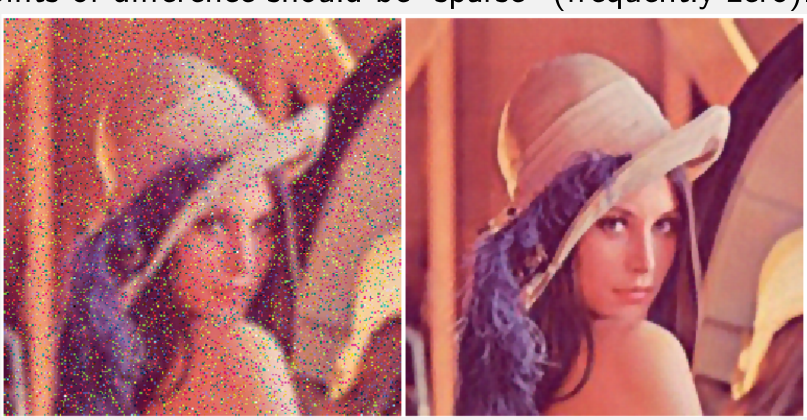
$$\min_{x_1,\dots,x_m} \min_{w^1,\dots,w^k \in \mathbb{R}^n} \sum_{j=1}^{k} \frac{1}{m} \sum_{i=1}^{m} \ell(y_i^k, (w^k)^\intercal x_i) + \lambda \Omega(w^k), \tag{15.116}$$

Norms, sparse norms, and computer vision

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

$$\Omega(w) = \sum_{i=2}^{N} |w_i - w_{i-1}| \tag{15.117}$$

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

Lovász extension and norms

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