CMSC 510 – L15
Regularization Methods for Machine Learning

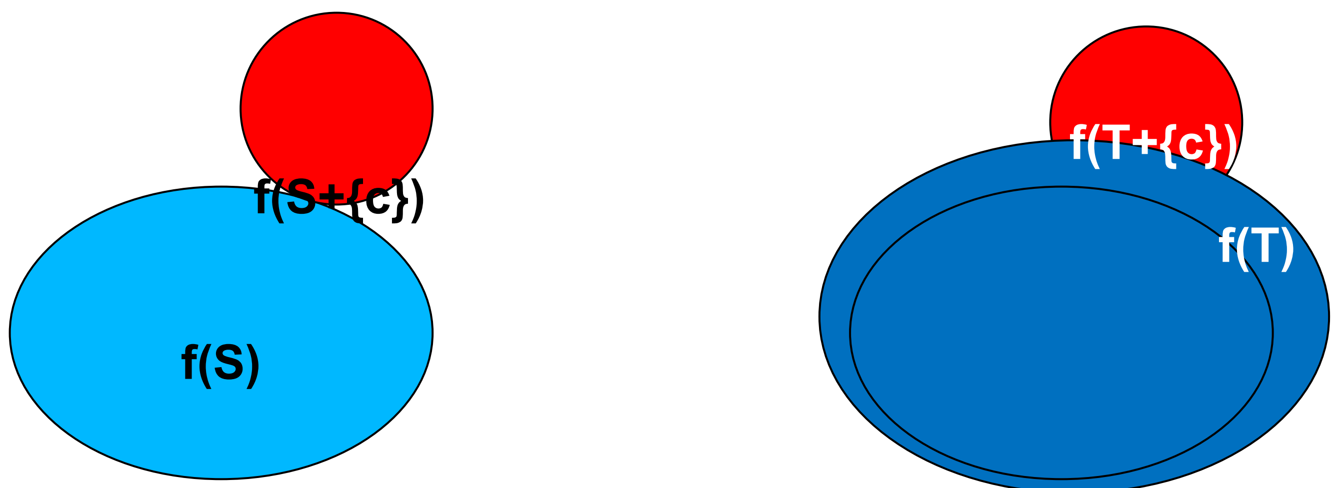
Instructor:
Dr. Tom Arodz

Recap: submodularity

- Set function over universe V, ie. Ω: 2^V -> R
- We focus on a family of set functions called **submodular set functions**

$$\forall S, T, \{c\} \subseteq V,\ \ S \subseteq T,\\ \Omega(S \cup \{c\}) - \Omega(S) \geq \Omega(T \cup \{c\}) - \Omega(T)$$

f(red+blue) - f(blue) >= f(red+navy) - f(navy)



- Adding the same element {c} to a set T increases Ω less then adding it to a subset S of T

Alternative definitions

- **Modular** set functions: $\Omega(\{a,b,\ldots,z\})=\Omega_a+\Omega_b+\ldots+\Omega_z$

$$\forall A,B \subseteq V \quad \Omega(A \cup B) + \Omega(A \cap B) = \Omega(A) + \Omega(B)$$

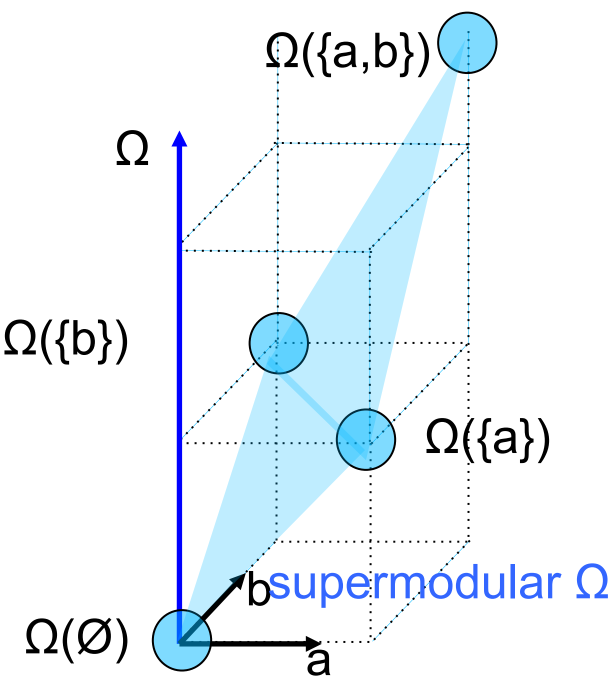
- **Sub**modular set functions: $\Omega(\{a,b,\ldots,z\}) \leq \Omega_a+\Omega_b+\ldots+\Omega_z$

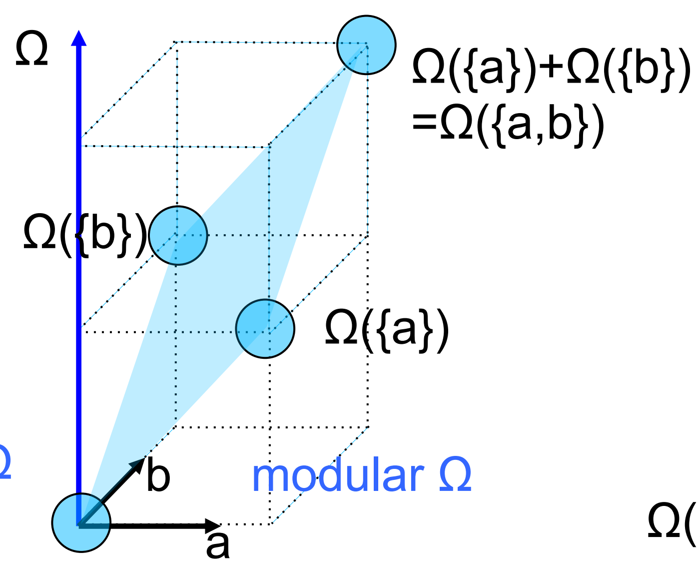
$$\Omega(A \cup B) + \Omega(A \cap B) \leq \Omega(A) + \Omega(B)$$

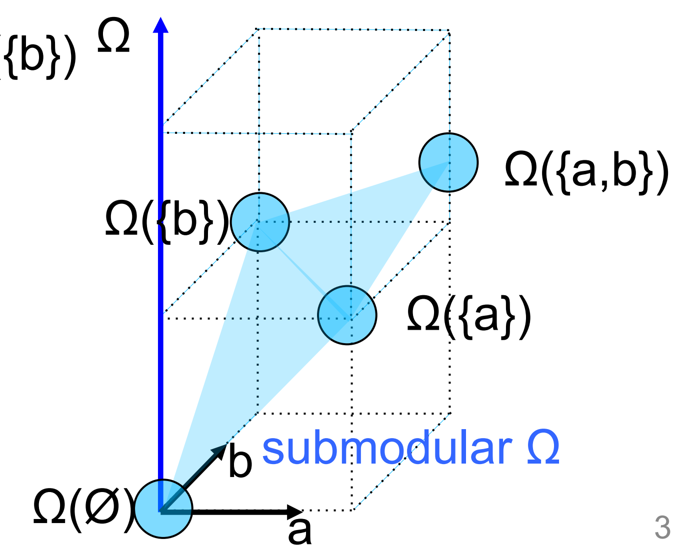
- **Super**modular set functions: $\Omega(\{a,b,\ldots,z\}) \geq \Omega_a+\Omega_b+\ldots+\Omega_z$

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

- If Ω is modular, it is both submodular and supermodular
  - If Ω is submodular, -Ω is supermodular

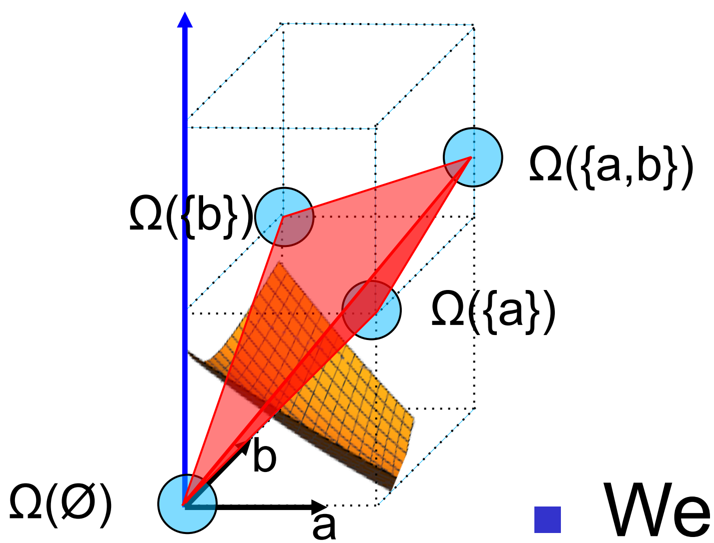


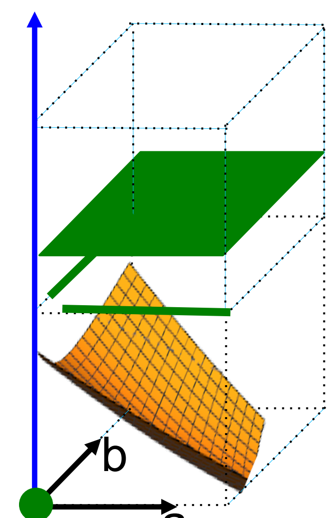




Recap: submodular regularizers

- Given a set function **$\Omega(S)=\Omega(w)$** where $S=1_w$
- We have a well-defined Lovasz extension Ω^L: $[0,1]^n \to R$





- We can use
  $\text{argmin}_w\, R_S(w)+\Omega^L(w)$
  - we'll need proximal descent
- Instead:
  $\text{argmin}_w\, R_S(w)+\Omega([w])$

Big picture

- We can solve problems of the form:
  - **Differentiable risk**
    + convex extension of a submodular set function Ω

- What are some interesting submodular set functions?
  - What are their extensions?
  - Can we solve proximal operator for them?

Submodular set functions

Assume w.l.o.g. Ω(Ø)=0

- **Convex closure of Ω: Ω⁻**

  Pointwise highest convex function that bounds Ω from below

  Complicated to evaluate what the function **Ω⁻(w)** actually is.

- **Lovasz extension of Ω: Ω^L**

$$\Omega^L(w) = \sum_{i=0}^{F} \lambda_i \Omega(S_i) \qquad F = |V|$$

where S_i is a chain of nested sets:

$$\emptyset = S_0 \subset S_1 \subset S_2 \subset \ldots \subset S_F = V$$

such that:

$$\sum_{i=0}^{F} \lambda_i 1_{S_i} = w, \sum_{i=0}^{F} \lambda_i = 1, \lambda_i \geq 0$$

Submodular set functions

$F = |V|$

- We'll see a condition like this:

$$\emptyset = S_0 \subset S_1 \subset S_2 \subset \ldots \subset S_F = V$$

$$\sum_{i=0}^{F} \lambda_i 1_{S_i} = w, \sum_{i=0}^{F} \lambda_i = 1, \lambda_i \geq 0$$

S_0					λ_0
S_1		λ_1			
S_2		λ_2		λ_2	
S_3	λ_3	λ_3		λ_3	
S_4	λ_4	λ_4	λ_4	λ_4	
	w_1	w_2	w_3	w_4	

Submodular set functions

Assume w.l.o.g. $\Omega(\emptyset)=0$

- **Convex closure of Ω: Ω^-**

  Pointwise highest convex function that bounds Ω from below

  Complicated to evaluate what the function $\mathbf{\Omega^-(w)}$ actually is.

- **Lovasz extension of Ω: Ω^L**

$$\Omega^L(w) = \sum_{i=0}^{F} \lambda_i \Omega(S_i) \qquad F = |V|$$

where S_i is a chain of nested sets:

$$\emptyset = S_0 \subset S_1 \subset S_2 \subset \ldots \subset S_F = V$$

such that:

$$\sum_{i=0}^{F} \lambda_i 1_{S_i} = w, \sum_{i=0}^{F} \lambda_i = 1, \lambda_i \geq 0$$

For any feature **i**, total weight on $\mathbf{v_i}$ (from sum of $\lambda_j 1_{S_j}$) is equal to $\mathbf{w_i}$

- Lovasz extension may not be convex for arbitrary set functions
- But for submodular set functions, $\mathbf{\Omega^L(w)=\Omega^-(w)}$, it is convex
  - The converse is also true, if Ω^L is convex then Ω is submodular

Submodular set functions

- **Lovasz extension of Ω: Ω^L** $\Omega^L(w) = \sum_{i=0}^{F} \lambda_i \Omega(S_i)$

  where: $\emptyset = S_0 \subset S_1 \subset S_2 \subset \ldots \subset S_F = V$

  $\sum_{i=0}^{F} \lambda_i 1_{S_i} = w, \sum_{i=0}^{F} \lambda_i = 1, \lambda_i \geq 0$
- **How to evaluate it for vector w in $[0,1]^F$?**
  - What are the λ_i and S_i
  - Order elements in V in decreasing order of w's:
    - $V=\{v_1,v_2,\ldots,v_F\}$ such that $w_1 \geq w_2 \geq \ldots \geq w_F \geq 0$
  - Set: $S_0=\emptyset$, $\lambda_0=1-w_1$

    $S_1=\{v_1\}$ $\lambda_1=w_1 - w_2$

    $S_2=\{v_1,v_2\}$, $\lambda_2=w_2 - w_3$

    …

    $S_i=S_{i-1}+\{v_i\}=\{v_1,\ldots,v_i\}$, $\lambda_i = w_i - w_{i+1}$

    …

    $S_F=V$ $\lambda_F=w_F$

Submodular set functions

S_0					λ_0
S_1	λ_1				
S_2	λ_2	λ_2			
S_3	λ_3	λ_3	λ_3		
S_4	λ_4	λ_4	λ_4	λ_4	
	w_1	w_2	w_3	w_4	

- $V=\{v_1,v_2,\dots,v_F\}$ such that $w_1 \geq w_2 \geq \dots \geq w_F \geq 0$
- Set: $S_0=\emptyset$, $\lambda_0=1-w_1$
  $S_1=\{v_1\}$ $\lambda_1=w_1-w_2$
  $S_2=\{v_1,v_2\}$, $\lambda_2=w_2-w_3$
  $S_3=\{v_1,v_2,v_3\}$, $\lambda_3=w_3-w_4$
  $S_4=V$ $\lambda_4=w_4$

$$\sum_{i=0}^{F}\lambda_i 1_{S_i}=w, \sum_{i=0}^{F}\lambda_i=1, \lambda_i\geq 0$$

Submodular set functions

- **Lovasz extension of Ω: Ω^L** $\Omega^L(w) = \sum_{i=0}^{F} \lambda_i \Omega(S_i)$

  where: $\emptyset = S_0 \subset S_1 \subset S_2 \subset \ldots \subset S_F = V$

  $\sum_{i=0}^{F} \lambda_i 1_{S_i} = w, \sum_{i=0}^{F} \lambda_i = 1, \lambda_i \geq 0$

- **How to evaluate it for vector w in $[0,1]^F$?**
  - Order elements in V in decreasing order of w's:
    - $V=\{v_1,v_2,\ldots,v_F\}$ such that $w_1 \geq w_2 \geq \ldots \geq w_F \geq 0$
  - Set:
    $S_0=\emptyset$, $\lambda_0=1-w_1$
    $S_i=S_{i-1}+\{v_i\}=\{v_1,\ldots,v_i\}$, $\lambda_i = w_i - w_{i+1}$
    $S_F=V$ $\lambda_F=w_F$

- This set up meets the conditions on S_i , λ_i :
  - S_i are nested, $\lambda_i \geq 0$ and add up to 1
  - For any feature **i**, total weight on $\mathbf{v_i}$ (from sum of $\lambda_j 1_{S_j}$) is equal to $\mathbf{w_i}$

$$\sum_{S_j : v_i \in S_j} \lambda_j = \sum_{S_j : j \geq i} \lambda_j = \sum_{j=i}^{F} \lambda_j =$$

$$= w_F + \sum_{j=i}^{F-1} w_j - w_{j+1} = w_i + \sum_{j=i+1}^{F} w_j - w_j = w_i$$

Submodular set functions

- **Lovasz extension of Ω: Ω^L** $\Omega^L(w) = \sum_{i=0}^{F} \lambda_i \Omega(S_i)$

  where: $\emptyset = S_0 \subset S_1 \subset S_2 \subset \ldots \subset S_F = V$

  $\sum_{i=0}^{F} \lambda_i 1_{S_i} = w, \sum_{i=0}^{F} \lambda_i = 1, \lambda_i \geq 0$
- **How to evaluate it for vector w in $[0,1]^F$?**
  - Order elements in V in decreasing order of w's:
    - $V=\{v_1,v_2,\ldots,v_F\}$ such that $w_1 \geq w_2 \geq \ldots \geq w_F$
  - Set: $S_0=\emptyset$, $\lambda_0=1-w_1$

    $S_i=S_{i-1}+\{v_i\}=\{v_1,\ldots,v_i\}$, $\lambda_i = w_i - w_{i+1}$

    $S_F=V$ $\lambda_F=w_F$
  - S_i and λ_i depend fully on vector w
    - For a given w, these are easy to obtain!
  - But S_i and λ_i do not depend at all on function **Ω**
    - No need to consider different sets of λ_i etc.

Lovasz extension $\Omega^L(w) = \sum_{i=0}^F \lambda_i \Omega(S_i)$

$V=\{v_1,v_2,\ldots,v_F\}$ such that $w_1 \geq w_2 \geq \ldots \geq w_F \geq 0$

$S_0=\emptyset$, $\lambda_0=1-w_1$

$S_i=S_{i-1}+\{v_i\}=\{v_1,\ldots,v_i\}$, $\lambda_i = w_i - w_{i+1}$

$S_F=V$ $\lambda_F=w_F$

- **It might be easier to work with a restructured formula:**

$$\Omega^L(w) = \sum_{i=0}^{F} \lambda_i \Omega(S_i) = \lambda_0 \Omega(\emptyset) + \lambda_F \Omega(S_F) + \sum_{i=1}^{F-1} \lambda_i \Omega(S_i)$$

$$= (1-w_1)\Omega(S_0) + w_F \Omega(S_F) + \sum_{i=1}^{F-1} (w_i - w_{i+1}) \Omega(S_i)$$

$$= (1\Omega(S_0) - w_1 \Omega(S_0) + w_F \Omega(S_F) + \sum_{i=1}^{F-1} w_i \Omega(S_i) - \sum_{i=2}^{F} w_i \Omega(S_{i-1})$$

$$= \Omega(S_0) + \sum_{i=1}^{F} w_i \Omega(S_i) - \sum_{i=1}^{F} w_i \Omega(S_{i-1})$$

$$= \boxed{\Omega(S_0) + \sum_{i=1}^{F} w_i \left[\Omega(S_i) - \Omega(S_{i-1})\right]}$$

Submodular set functions

- **Lovasz extension of Ω: Ω^L** $\Omega^L(w) = \sum_{i=0}^{F} \lambda_i \Omega(S_i)$

  where: $\emptyset = S_0 \subset S_1 \subset S_2 \subset \ldots \subset S_F = V$

  $\sum_{i=0}^{F} \lambda_i 1_{S_i} = w, \sum_{i=0}^{F} \lambda_i = 1, \lambda_i \geq 0$
- **How to evaluate it for vector w in $[0,1]^F$?**
  - Order elements in V in decreasing order of w's:
    - $V=\{v_1,v_2,\ldots,v_F\}$ such that $w_1 \geq w_2 \geq \ldots \geq w_F$
  - Set: $S_0=\emptyset$, $\lambda_0=1-w_1$

    $S_i=S_{i-1}+\{v_i\}=\{v_1,\ldots,v_i\}$, $\lambda_i = w_i - w_{i+1}$

    $S_F=V$ $\lambda_F=w_F$
  - Two alternative formulas:

$$\Omega^L(w) = (1-w_1)\Omega(S_0) + w_F \Omega(S_F) + \sum_{i=1}^{F-1} (w_i - w_{i+1}) \Omega(S_i)$$

$$= \Omega(S_0) + \sum_{i=1}^{F} w_i \left[\Omega(S_i) - \Omega(S_{i-1})\right]$$

Lovasz extension

$V=\{v_1,v_2,\ldots,v_F\}$ such that $w_1 \geq w_2 \geq \ldots \geq w_F \geq 0$

$S_0=\emptyset$, $\lambda_0=1-w_1$
$S_i=S_{i-1}+\{v_i\}=\{v_1,\ldots,v_i\}$, $\lambda_i = w_i - w_{i+1}$
$S_F=V$ $\lambda_F=w_F$

- How to use it in practice?

- Example: modular function $\Omega(S) = \sum_{v_k \in S} \Omega(v_k)$ **e.g. $\Omega(v_k)=1$**

$$\Omega^L(w) = \Omega(S_0) + \sum_{i=1}^{F} w_i \left[\Omega(S_i) - \Omega(S_{i-1})\right]$$

$$= \Omega(\emptyset) + \sum_{i=1}^{F} w_i \Omega(v_i)$$

We now have an analytical formula for Ω^L

$$= \sum_{i=1}^{F} w_i \Omega(v_i)$$

Lovasz extension

- How to use it in practice? Example: modular function $\Omega(S) = \sum_{v_k \in S} \Omega(v_k)$
  - Special case: $\Omega(v_i) = \frac{1}{C}$
- Lovasz extension (and thus our formula) is defined over unit cube:

$$\Omega^L : [0,1]^F \to \mathbb{R}$$

$$\Omega^L(w) = \frac{1}{C} \sum_{i=1}^{F} w_i = \frac{1}{C} \sum_{i=1}^{F} |w_i| = \frac{1}{C} ||w||_1$$

- Let's define a function defined on the whole range (from – to + infinity)

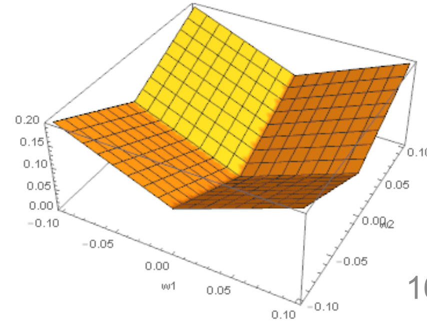
$$\Omega^{L\infty} : \mathbb{R}^F \to \mathbb{R}$$

$$\Omega^{L\infty}(w) = \frac{1}{C} \sum_{i=1}^{F} |w_i| = \frac{1}{C} ||w||_1$$

To simplify notation, often we'll just treat $\mathbf{\Omega}^L$ as if it was defined over whole R^F

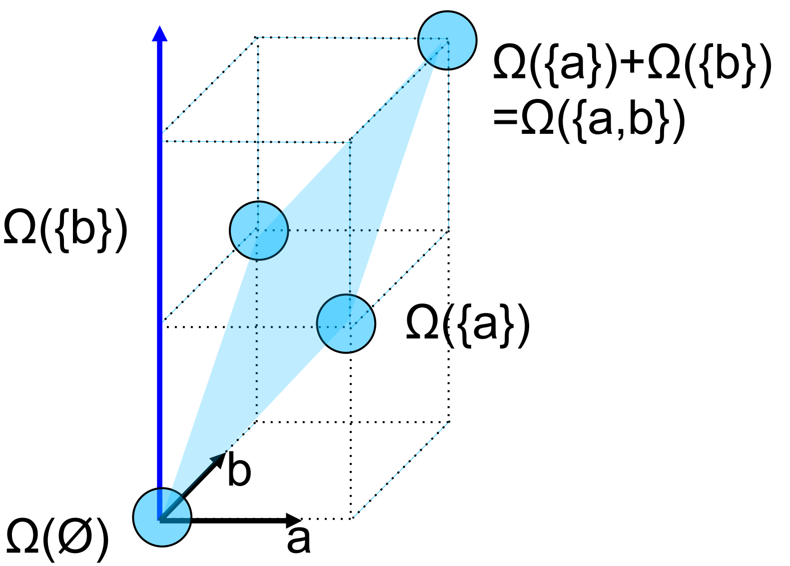
- We can use that formula as penalty term: $\arg\min_w \frac{1}{C} \|w\|_p^p + \sum^m \ln(1 + \mathrm{e}^{-y_i w^T x_i})$

- L_1 penalty ⇔ special case of Lovasz extension of modular set function defined over sets of features

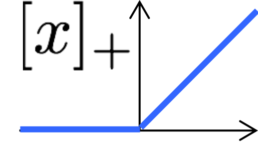


Set cardinality

- **Set cardinality** is a modular (and thus submodular) set function
  - $\Omega(S)=|S|$
- Interpretation of Ω in machine learning:
  - $\Omega([w])$ = number of features in the model represented by vector w
  - Model preferred by $\Omega^L(w)$: lost of feature weights w_f are 0
- Lovasz extension on $[0,1]^F$:
  - $\Omega^L(w) = \Sigma_f w_f$ (we derived it previously)
  - We have derived it from definition of Ω^L
- Extension to R^F:
  - $\Omega^{L\infty}(w) = ||w||_1 = \Sigma_f |w_f|$
  - L_1 penalty
- Minimum of $prox_{\Omega^L}(v)$:
  - soft thresholding of v



Graph cut capacity



$$[x]_+ = \max(x, 0)$$

- **Directed graph cut capacity** is a submodular set function
  - Ω(S) = total weights of edges **from S to V-S**

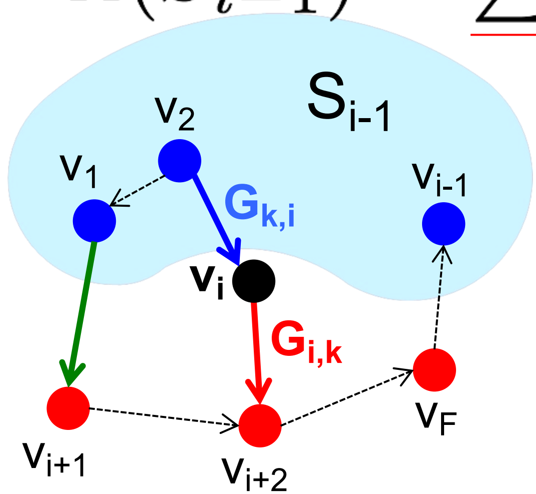
$$\Omega(S) = \sum_{j \in S} \sum_{k \notin S} G_{jk}$$

$$\Omega^L(w) = \sum_{j=1}^{F} \sum_{k=1}^{F} G_{j,k} \left[w_j - w_k\right]_+$$

- Formula is: $\Omega^L(w) = \sum_{i=1}^{F} w_i \left[\Omega(S_i) - \Omega(S_{i-1})\right]$
  - Where we ordered the vertices in decreasing order of w_i

$$\Omega(S_i) - \Omega(S_{i-1}) = \underbrace{\sum_{k=i+1}^{F} G_{i,k}}_{\text{red edges}} - \underbrace{\sum_{k=1}^{i-1} G_{k,i}}_{\text{blue edges}}$$

$S_i = S_{i-1} + \{v_i\}$
$= \{v_1, \ldots, v_i\}$



Move from S_{i-1} to S_i

Red edges start playing a role

Blue edges stop playing a role

Green edges: no change

Black edges: play no role

Graph cut capacity

- Starting point:

$$\Omega^L(w) = \sum_{i=1}^{F} w_i \left[\Omega(S_i) - \Omega(S_{i-1})\right]$$

$$\Omega(S_i) - \Omega(S_{i-1}) = \sum_{k=i+1}^{F} G_{i,k} - \sum_{k=1}^{i-1} G_{k,i}$$

- So:

$$\Omega^L(w) = \sum_{i=1}^{F} w_i \sum_{k=i+1}^{F} G_{i,k} - \sum_{i=1}^{F} w_i \sum_{k=1}^{i-1} G_{k,i}$$

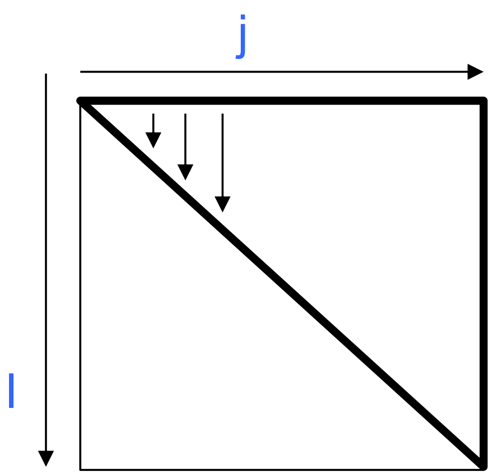
Graph cut capacity

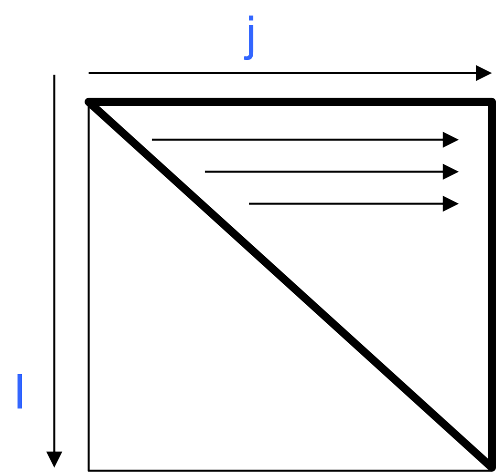
- Formula up to now:

$$\Omega^L(w) = \sum_{i=1}^{F} w_i \sum_{k=i+1}^{F} G_{i,k} - \sum_{i=1}^{F} w_i \sum_{k=1}^{i-1} G_{k,i}$$

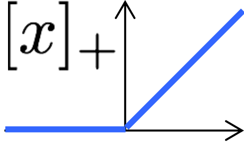
- Reorder the two sums, change variable names, first change: i->j, k->l then change j->k, l->i :

$$- \sum_{j=1}^{F} w_j \sum_{l=1}^{j-1} G_{l,j} = \sum_{l=1}^{F} \sum_{j=l+1}^{F} -w_j G_{l,j} = \sum_{i=1}^{F} \sum_{k=i+1}^{F} -w_k G_{i,k}$$





Graph cut capacity



$$[x]_+ = \max(x, 0)$$

- Formula up to now:

$$\Omega^L(w) = \sum_{i=1}^{F} w_i \sum_{k=i+1}^{F} G_{i,k} - \sum_{i=1}^{F} w_i \sum_{k=1}^{i-1} G_{k,i}$$

- Reorder the 2nd sum, change variable names (i->j->k, k->l->i):

$$-\sum_{j=1}^{F} w_j \sum_{l=1}^{j-1} G_{l,j} = \sum_{l=1}^{F} \sum_{j=l+1}^{F} -w_j G_{l,j} = \sum_{i=1}^{F} \sum_{k=i+1}^{F} -w_k G_{i,k}$$

- We get:

$$\Omega^L(w) = \sum_{i=1}^{F} \sum_{k=i+1}^{F} G_{i,k}(w_i - w_k)$$

- We ordered our vertices in decreasing way, so:

$$k < i \implies w_i - w_k \leq 0$$

- We can expand the sum and get the **final formula**:

$$\Omega^L(w) = \sum_{i=1}^{F} \sum_{k=1}^{F} G_{i,k}[w_i - w_k]_+$$