Combinatorial Optimization in Computer Vision (IN2245)

Frank R. Schmidt
Csaba Domokos

Winter Semester 2015/2016

Pseudo-Boolean Function

Boolean Variables

A Boolean variable $x \in \mathbb{B}$ can either be *true* or *false*.

To simplify the notation, we denote the Boolean set as $\mathbb{B} := \{0, 1\}$.
Here, 0 and 1 are identified with *false* and *true* respectively.

$\mathbb{B}$ forms a **totally ordered set**,*i.e.*,

$$x \leqslant y \text{ and } y \leqslant x \Leftrightarrow x = y \qquad \text{(for all } x, y \in \mathbb{B})$$
$$x \leqslant y \text{ and } y \leqslant z \Rightarrow x \leqslant z \qquad \text{(for all } x, y, z \in \mathbb{B})$$
$$x \leqslant y \text{ or } y \leqslant x \qquad \text{(for all } x, y \in \mathbb{B})$$

For two Boolean variables $x, y \in \mathbb{B}$, we denote

$$x \wedge y := \min\{x, y\} \qquad x \vee y := \max\{x, y\}$$

Powerset

Given an arbitrary set Ω, we denote the *powerset* of Ω as $\mathcal{P}(\Omega)$ or 2^Ω.
The powerset is the unique set that contains all subsets of Ω.

For two sets $A, B \in \mathcal{P}(\Omega)$, the subset relationship

$$A \subset B :\Leftrightarrow [\forall i \in A : i \in B]$$

makes $\mathcal{P}(\Omega)$ a **partially ordered set**,*i.e.*,

$$A \subset B \text{ and } B \subset A \Leftrightarrow A = B \qquad \text{(for all } A, B \in \mathcal{P}(\Omega))$$

$$A \subset B \text{ and } B \subset C \Rightarrow A \subset C \qquad \text{(for all } A, B, C \in \mathcal{P}(\Omega))$$

For two subsets $A, B \in \mathcal{P}(\Omega)$, we denote

$$A \cap B := \text{meet}(A, B) \qquad A \cup B := \text{join}(A, B)$$

Union and Intersection

Note that for $A, B \in \mathcal{P}(\Omega)$, meet$(A, B)$ is defined as the *maximal lower bound* of A and B,*i.e.*, meet(A, B) is the $C \in \mathcal{P}(\Omega)$ such that

- C is a lower bound,*i.e.*, $C \subset A$ and $C \subset B$.
- For all other lower bounds D, $D \subset C$ holds.

One can show that $\cap$ and $\cup$ coincides with the classial notion of *union* and *intersection*:

$$A \cup B = \{i \in \Omega | i \in A \text{ or } i \in B\}$$
$$A \cap B = \{i \in \Omega | i \in A \text{ and } i \in B\}$$

Subsets as Boolean Mappings

To each subset $A \in \mathcal{P}(\Omega)$, we can define the characteristic function

$$\chi_A : \Omega \to \mathbb{B}$$
$$i \mapsto [i \in A]$$

For two characteristic functions χ_A and χ_B, we can define

$$[\chi_A \wedge \chi_B](i) := \chi_A(i) \wedge \chi_B(i) \qquad [\chi_A \vee \chi_B](i) := \chi_A(i) \vee \chi_B(i)$$

and we obtain

$$\chi_A \wedge \chi_B = \chi_{A \cap B} \qquad \chi_A \vee \chi_B = \chi_{A \cup B}$$

The partial ordering of $\mathcal{P}(\Omega)$ is induced by the total ordering of $\mathbb{B}$.

If we replace $\mathbb{B}$ with a totally ordered set $\mathcal{L}$, wie can replace $\mathcal{P}(\Omega)$ with $\mathcal{L}^\Omega$.

Boolean Function

A **Boolean function** $E : 2^\Omega \to \mathbb{B}$ assigns to every subset $A \subset \Omega$ a Boolean value $E(A)$.

One can use a Boolean function in order to test certain properties:

$$\begin{aligned} E_1(A) =& [A \neq \varnothing] \\ E_2(A) =& [A \text{ is connected}] \\ E_3(A) =& [A \text{ is a square}] \\ E_4(A) =& [A \text{ is almost circular}] \end{aligned}$$

In Computer Vision, we are usually interested in problems that are beyond a pure *satisfiability test*.

We are not interested whether A is almost circular. Instead, we would like to evaluate some sort of dissimilarity measure between A and a perfect disc.

Pseudo-Boolean Function

A **pseudo-Boolean function** $E : 2^\Omega \to \mathbb{R}$ assigns to every subset $A \subset \Omega$ a real value $E(A)$.

In the following, we will identify a subset $A \subset \Omega$ with its characteristic function $\chi_A : \Omega \to \mathbb{B}$. For disjoint sets A and B, we will write $A + B := A \cup B$ and for subsets $S \subset T$, we will write $T - S := T \backslash S$.

Since sets are identified with binary functions, we may also refer to E as a functional. In the literature, one usually talks about E as a function if Ω is a finite set. E is referred to as a functional if Ω is a continuous set
(real-valued vector spaces, finite-dimensional manifolds, etc.).

In this lecture, we will only consider finite sets Ω.
See Variational Methods for Computer Vision for functional-driven optimization methods.

Pseudo-Boolean Optimization

Most Computer Vision problems can be cast as the minimization of a pseudo-Boolean function $E : 2^\Omega \to \mathbb{R}$.

Given $E : 2^\Omega \to \mathbb{R}$, we are interested in the global minimum $\min_{A \subset \Omega} E(A)$ and in one of its global minimizers $A \in \operatorname{argmin} E$,

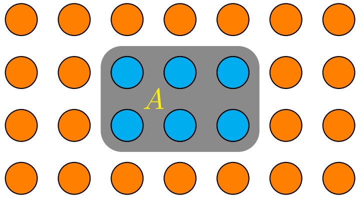
$$\operatorname{argmin} E := \{A \subset \Omega | E(A) \leqslant E(B) \text{ for all } B \subset \Omega\}.$$

Since Ω is finite, we know that $\operatorname{argmin} E$ is not empty, but in general it may contain more than one global minimizer.

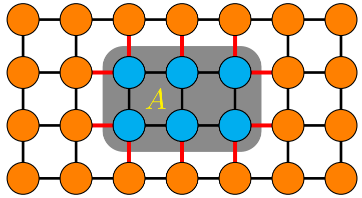
If the computation of a global minimizer is NP-hard, we are also satisfied with an approximation. A set $S \subset \Omega$ is called an $(1 + \epsilon)$-approximation of $\operatorname{argmin} E$, if the following holds

$$E(S) \leqslant (1 + \epsilon) \cdot \min_{A \subset \Omega} E(A).$$

Binary Image Segmentation



Data Term



Length Term

Segmenting an image can be cast as minimizing the energy

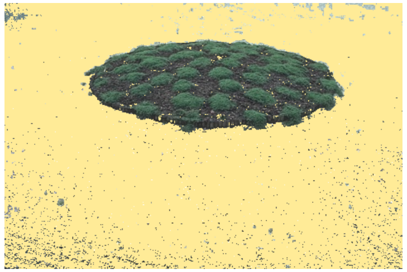
$$E_{\text{Data}}(A) = \sum_{i \in A} f(i)$$

It is common to combine it with a length term

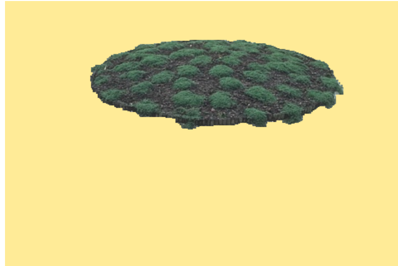
$$E_{\text{Length}}(A) = \sum_{i \in A} \sum_{\substack{j \notin A, \\ |i-j|=1}} 1$$

Binary Image Segmentation

Given Image



Minimizing Data Term



Minimizing Data + Length Term

$$
\begin{aligned}
\underset{A\subset\Omega}{\operatorname{argmin}}\, E(A) &= \underset{A\subset\Omega}{\operatorname{argmin}} \sum_{i\in\Omega-A} f_0(i) + \sum_{i\in A} f_1(i) + \text{length}(A) \\
&= \underset{A\subset\Omega}{\operatorname{argmin}} \sum_{i\in\Omega} f_0(i) + \sum_{i\in A} [\underbrace{f_1(i) - f_0(i)}_{=:f(i)}] + \text{length}(A) \\
&= \underset{A\subset\Omega}{\operatorname{argmin}} \sum_{i\in A} f(i) + \text{length}(A)
\end{aligned}
$$

We will show that this energy can be minimized in polynomial time.

Modular Functions

The minimization of a pseudo-Boolean function $E\colon 2^\Omega \to \mathbb{R}$ (with $E(\emptyset) = 0$) becomes very easy, if E is modular,*i.e.*,

$$E(A \cup B) + E(A \cap B) = E(A) + E(B) \qquad \text{(for all } A, B \in 2^\Omega\text{)}$$

For disjoint $A, B \in 2^\Omega$, we have $E(A + B) = E(A) + E(B)$, which implies

$$E(A) = \sum_{i \in A} E(\{i\}).$$

A global minimizer of the modular function E is therefore

$$A = \{i \in \Omega | E(\{i\}) < 0\}$$

and it can be found in $\mathcal{O}(N)$ time, where $N := |\Omega|$ is the cardinality of Ω.

Submodularity and Supermodularity

A pseudo-Boolean function $E\colon 2^\Omega \to \mathbb{R}$ is called submodular if

$$E(A \cup B) + E(A \cap B) \leqslant E(A) + E(B) \qquad \text{(for all } A, B \in 2^\Omega)$$

A pseudo-Boolean function $E\colon 2^\Omega \to \mathbb{R}$ is called supermodular if

$$E(A \cup B) + E(A \cap B) \geqslant E(A) + E(B) \qquad \text{(for all } A, B \in 2^\Omega)$$

Modular functions are submodular and supermodular. Modular, sub- and supermodular functions are closed w.r.t. summation and positive scaling.

Minimizing an arbitrary submodular functions can be done in polynomial time [Grötschel, Lovász, Schrijver, 1981].

The minimization of supermodular functions is NP-hard.

Submodularity w.r.t. 2 Variables

Let $E\colon 2^\Omega \to \mathbb{R}$ be submodular and let $S \in 2^\Omega$ and $i, j \in \Omega - S$. Then

$$E(S + \{i, j\}) + E(S) \leqslant E(S + \{i\}) + E(S + \{j\}). \qquad (1)$$

If we define $E_2\colon \mathbb{B} \times \mathbb{B} \to \mathbb{R}$ via $E_2(b_1, b_2) := E(S + b_1 \cdot \{i\} + b_2 \cdot \{j\})$, we can rewrite (1) as

$$E_2(1, 1) + E_2(0, 0) \leqslant E_2(1, 0) + E_2(0, 1) \qquad (2)$$

If for a pseudo-Boolean function $E\colon 2^\Omega \to \mathbb{R}$, the Equation (1) is satisfied for all S, i, j, the energy E is in fact submodular. Some authors use therefore (2) as definition for submodularity.

Submodular Functions

$E_{\text{Length}} \colon 2^\Omega \to \mathbb{R}$ is submodular and $E_{\text{Data}} \colon 2^\Omega \to \mathbb{R}$ is modular.

Iff $E \colon 2^\Omega \to \mathbb{R}$ is a supermodular function, then $-E \colon 2^\Omega \to \mathbb{R}$ is submodular.

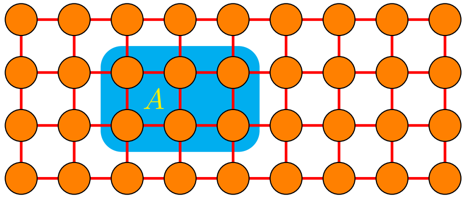
If $E \colon 2^\Omega \to \mathbb{R}$ is submodular, $T \subset \Omega$, then $E|T \colon 2^\Omega \to \mathbb{R}$ is submodular with

$$E|T(A) := E(T \cap A).$$

If $H \colon \mathbb{R} \to \mathbb{R}$ is a concave function, then $E_H \colon 2^\Omega \to \mathbb{R}$ is submodular with

$$E_H(A) := H(|A|).$$

Weighted Contour Length



Weighted Contour Length ($w < 0$)

The weighted contour length with negative weights is a supermodular energy.

Minimizing the length is equivalent to
maximizing the cut with positive weights.

The **Maximum Cut** problem is NP hard.

Thus, minimizing a supermodular function is in general NP hard.

Convex Closure

In order to analyze a pseudo-Boolean function $E\colon \mathbb{B}^\Omega \to \mathbb{R}$, one can extend it to a function $\bar{E}\colon [0,1]^\Omega \to \mathbb{R}$ such that $\bar{E}|\mathbb{B}^\Omega = E$.

Using a specific total ordering $\prec$ of the $N \in \mathbb{N}$ elements in Ω

$$i_1 \prec i_2 \prec \ldots \prec i_N,$$

we can write $E : \mathbb{B}^N \to \mathbb{R}$ and $\bar{E}\colon [0,1]^N \to \mathbb{R}$.

The **convex closure** $E^-\colon [0,1]^N \to \mathbb{R}$ is defined as

$$E^-(x) = \min\left\{ \sum_{S\subset\Omega} \alpha_S \cdot E(S) \middle| x = \sum_{S\subset\Omega} \alpha_S \cdot S, \sum_{S\subset\Omega} \alpha_S = 1, \alpha_S \geqslant 0 \right\}.$$

Note that E^- is piecewise linear and hence non-differentiable.

Convex Closure

Theorem 1. *Convex closure E^- of a pseudo-Boolean function E is convex.*

Proof. Let $x^0, x^1 \in [0,1]^N$, $\lambda \in [0,1]$ and $x^\lambda := (1-\lambda)\cdot x^0 + \lambda \cdot x^1$. We have to show that $E^-(x^\lambda) \leqslant (1-\lambda)E^-(x^0) + \lambda E^-(x^1)$. We have

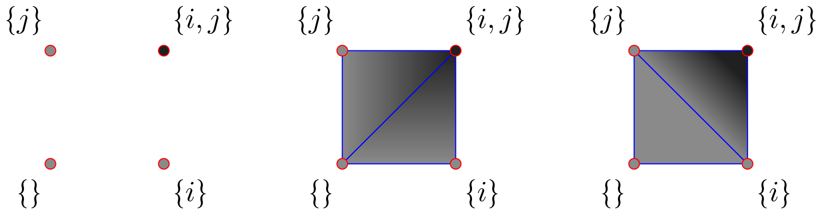
$$E^-(x^0) = \sum_{S\subset\Omega} \alpha_S^0 \cdot E(S) \qquad x^0 = \sum_{S\subset\Omega} \alpha_S^0 \cdot S$$

$$E^-(x^1) = \sum_{S\subset\Omega} \alpha_S^1 \cdot E(S) \qquad x^1 = \sum_{S\subset\Omega} \alpha_S^1 \cdot S$$

Defining $\alpha_S^\lambda := (1-\lambda)\cdot\alpha_S^0 + \lambda\cdot\alpha_S^1$, we obtain

$$E^-(x^\lambda) \leqslant \sum_{S\subset\Omega} \alpha_S^\lambda \cdot E(S) = (1-\lambda)E^-(x^0) + \lambda E^-(x^1)$$

Convex Closure (N=2)



Discrete Energy E | E^- for submodular E | E^- for supermodular E

Assume, we have $\Omega = \{i, j\}$ and the pseudo-Boolean function $E : 2^\Omega \to \mathbb{R}$

$$E(\emptyset) = E(\{i\}) = E(\{j\}) = 0 \qquad E(\{i, j\}) = \alpha \in \mathbb{R}$$

E is submodular for $\alpha \leqslant 0$ and supermodular for $\alpha \geqslant 0$.

The convex extension E^- is different for $\alpha < 0$ resp. $\alpha > 0$.

Lovász Extension

In general, it may take exponential time in order to evaluate E^-.

The **Lovász extension** on the other hand can be computed in linear time

$$E^L(x) = \sum_{n=0}^{k} \alpha_n \cdot E(S_n) \qquad \text{for } x = \sum_{n=0}^{k} \alpha_n \cdot S_n$$

$$\sum_{n=0}^{k} \alpha_n = 1, \alpha_n > 0$$

$$\varnothing \subset S_0 \subsetneq \ldots \subsetneq S_k \subset \Omega$$

Example 1. Let $\Omega = \{i, j\}$, $E\colon \mathbb{B}^\Omega \to \mathbb{R}$ a pseudo-Boolean function and $f = (0.1, 0.6)$. Then we have

$$S_0 = \varnothing; \quad S_1 = \{j\}; \quad S_2 = \{i, j\}$$
$$E^L(x) = 0.4 \cdot E(S_0) + 0.5 \cdot E(S_1) + 0.1 \cdot E(S_2)$$

Lovász Extension (Representation)

Theorem 2. *Let $x \in [0,1]^N$. Then there is a $k \leqslant N$, a chain $\varnothing \subset S_0 \subsetneq \ldots \subsetneq S_k \subset \Omega$ and $\alpha_0, \ldots, \alpha_k > 0$ such that $\sum_{n=0}^{k} \alpha_n = 1$ and $x = \sum_{n=0}^{k} \alpha_n S_n$. This representation is unique.*

Proof. Induction over $|X|$, $X = \{n | x_n > 0\}$. We will prove $k = |X| \leqslant N$.

Base Case: Assume that $|X| = 0$.
$X = \varnothing$ implies $x = 0$. We have uniquely $k = 0$, $S_0 = \varnothing$ and $\alpha_0 = 1$.

Inductive Step: Assume the theorem is true for all x' with $|X'| < |X|$.
The biggest set S_k has to be $\{n | x_n > 0\}$ and we have to choose $\alpha_k = \min X$. Otherwise, x is not representable as a convex combination. Let now $x' := x - \alpha_k S_k$. For the set X', we have $|X'| \leqslant |X| - 1$.
Therefore, there exists a unique representation $x' = \sum_{n=0}^{k-1} \alpha'_n S'_n$.
Since $\max X' \leqslant 1 - \alpha_k$, we have $S'_0 = \varnothing$ and $\alpha'_0 \geqslant \alpha_k$. Setting $\alpha_0 = \alpha'_0 - \alpha_k$, $\alpha_n = \alpha'_n$ for $0 < n < k$ and $S_n = S'_n$ for $0 \leqslant n < k$ provides us with the unique representation for x.

Lovász Extension

Theorem 3. *A pseudo-Boolean function E is submodular iff $E^- = E^L$.*

Proof.

Case 1: E is not submodular.
Then, there exist $S \subset \Omega$ and $i, j \in \Omega - S$ such that

$$E(S + \{i, j\}) + E(S) > E(S + \{i\}) + E(S + \{j\})$$

If we choose $x = S + \frac{1}{2}\{i\} + \frac{1}{2}\{j\}$, we have

$$\begin{aligned} E^L(x) &= \frac{1}{2}\left(E(S + \{i, j\}) + E(S)\right) \\ E^-(x) &\leqslant \frac{1}{2}\left(E(S + \{i\}) + E(S + \{j\})\right) \end{aligned}$$

and therefore $E^L \neq E^-$.

Lovász Extension

Proof (cont.).

Case 2: E is submodular.
Let $x \in [0,1]^N$ with $|\Omega| = N$ and

$$\mathcal{A} = \left\{ (\alpha_S)_{S\subset\Omega} \middle| x = \sum_{S\subset\Omega} \alpha_S \cdot S, \sum_{S\subset\Omega} \alpha_S = 1, E^-(x) = \sum_{S\subset\Omega} \alpha_S E(S) \right\}.$$

We choose an $\boldsymbol{\alpha} \in \mathcal{A}$ that maximizes $\sum_{S\subset\Omega} \alpha_S \cdot |S|^2$. We have to prove that the α_S are only positive for sets that are subsets from one another. Assume that there are $S, T \subset \Omega$ with $\alpha_S \geqslant \alpha_T > 0$ and $|S\backslash T|, |T\backslash S| > 0$. Replacing $\alpha_T(S+T)$ with $\alpha_T(S\cap T + S\cup T)$ does not increase the energy due to submodularity, but

$$|S\cap T|^2 + |S\cup T|^2 = |S|^2 + |T|^2 + 2\,|S\backslash T| \cdot |T\backslash S| > |S|^2 + |T|^2,$$

which contradicts the choice of $\boldsymbol{\alpha}$.

Lovász Extension

For submodular functions E, we saw

1. The Lovász extension E^L can be evaluated in polynomial time.
2. Since $E^L = E^-$, we can minimize E^L in polynomial time.
3. Since E^L is piecewise linear, the minimum is been taken at its boundary. Therefore, the minimum of E^L is been taken by a set $S \subset \Omega$.

[Grötschel, Lovász, Schrijver: *The ellipsoid method and its consequences in combinatorial optimization*, Combinatorica 1 (1981)]

> "The algorithm [...] is based on the ellipsoid method, and uses therefore a heavy framework of division, rounding, and approximation; moreover, it is not practical."
>
> *A. Schrijver, 2000*

Schrijver's new method takes $\mathcal{O}(N^5)$ iterations. In each iteration, an $N \times N$ matrix has to be inverted.

Multilinear Extension

Another extension of a pseudo-Boolean function $E : \mathbb{B}^N \to \mathbb{R}$ is the multilinear extension $\bar{E} : [0,1]^N \to \mathbb{R}$.
It makes use of the fact that for a given set $A \subset \Omega$ the function

$$\begin{aligned} F : \mathbb{B}^N &\to \mathbb{R} \\ (x_1, \ldots, x_n) &\mapsto \prod_{i \in A} x_i \prod_{i \notin A} (1 - x_i) \end{aligned}$$

satisfies

$$F(S) = \begin{cases} 1 & \text{if } S = A \\ 0 & \text{otherwise} \end{cases}$$

The **multilinear extension** $\bar{E}$ is defined via

$$\bar{E}(x_1, \ldots, x_n) := \sum_{A \subset \Omega} E(A) \cdot \prod_{i \in A} x_i \prod_{i \notin A} (1 - x_i)$$

Multilinear Extension (Example)

Consider the pseudo-Boolean function $E : \mathbb{B}^3 \to \mathbb{R}$

x_1	x_2	x_3	$E(x_1, x_2, x_3)$
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	0
1	0	0	1
1	0	1	1
1	1	0	1
1	1	1	0

and its extension $\bar{E}\colon [0,1]^N \to \mathbb{R}$:

$$\bar{E}(x_1, x_2, x_3) = x_1(1 - x_2)(1 - x_3) + x_1(1 - x_2)x_3 + x_1x_2(1 - x_3).$$

Using the notation $\bar{x} := (1 - x)$, we can write $\bar{E}$ as

$$\begin{aligned}\bar{E}(x_1, x_2, x_3) =& x_1\bar{x}_2\bar{x}_3 + x_1\bar{x}_2x_3 + x_1x_2\bar{x}_3 \\ =& x_1\left(1 - x_2x_3\right)\end{aligned}$$

Second Derivatives

Theorem 4. *Iff E is submodular, we have $\frac{\partial^2 \bar{E}}{\partial x_i \partial x_j} \leqslant 0$ for all x_i, x_j.*

Proof. We have

$$\begin{aligned}
\frac{\partial \bar{E}}{\partial x_i} &= \sum_{A \subset \Omega} E(A) \frac{\partial}{\partial x_i} \left[\prod_{j \in A} x_j \prod_{j \notin A} \bar{x}_j \right] \\
&= \sum_{i \in A \subset \Omega} E(A) \left[\prod_{j \in A, j \neq i} x_j \prod_{j \notin A} \bar{x}_j \right] - \sum_{i \notin A \subset \Omega} E(A) \left[\prod_{j \in A} x_j \prod_{j \notin A, j \neq i} \bar{x}_j \right] \\
&= \bar{E}(x_1, \ldots, x_{i-1}, 1, x_{i+1}, \ldots, x_n) - \bar{E}(x_1, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_n) \\
&= \sum_{A \subset (\Omega \setminus \{i\})} [E(A+i) - E(A)] \left[\prod_{j \in A} x_j \prod_{j \notin A} \bar{x}_j \right]
\end{aligned}$$

Second Derivatives

Proof (Cont.). For the second derivatives we get

$$\begin{aligned}
\frac{\partial^2 \bar{E}}{\partial x_j \partial x_i} &= \frac{\partial}{\partial x_j} \sum_{A \subset (\Omega \setminus \{i\})} [E(A+i) - E(A)] \left[\prod_{k \in A} x_k \prod_{k \notin A} \bar{x}_k \right] \\
&= \sum_{A \subset (\Omega \setminus \{i,j\})} [(E(A+i+j) - E(A+j)) - (E(A+i) - E(A))] \cdot \\
&\qquad \cdot \left[\prod_{j \in A} x_j \prod_{j \notin A} \bar{x}_j \right]
\end{aligned}$$

It follows that E is submodular iff $\frac{\partial^2 \bar{E}}{\partial x_j \partial x_i} \leqslant 0$.

Different Representations

$E\colon \mathbb{B}^N \to \mathbb{R}$ can be uniquely written as a **multi-linear function**

$$E(x) = \sum_{i=1}^{K} c_i \cdot \prod_{j \in \mathcal{C}_i} x_j,$$

where $c_i \in \mathbb{R}$ and $\mathcal{C}_i \subset \Omega$. We call $\mathcal{C}_i$ a **clique**. If the multi-linear function only contains cliques of size $|\mathcal{C}_i| \leqslant 2$, we call it a **quadratic** function.

We refer to Ω as the set of variables. The set $\mathcal{L} = \{x | x \in \Omega\} \sqcup \{\overline{x} | x \in \Omega\}$ is called the set of **literals**. Any pseudo-Boolean function $E\colon \mathbb{B} \to \mathbb{R}$ can be written as a **posiform**

$$E(x) = \sum_{i=1}^{K} c_i \cdot \prod_{j \in \mathcal{C}_i} x_j + C_0,$$

where $c_i > 0$, $C_0 \in \mathbb{R}$ and $\mathcal{C}_i \subset \mathcal{L}$. This representation is **not** unique.

Literature

Pseudo Boolean Optimization

- Boros and Hammer, *Pseudo-Boolean Optimization*, 2002, Discrete Applied Mathematics (123), 155–225.

Submodularity

- Edmonds, *Submodular Functions, Matroids, and Certain Polyhedra*, 1970, Combinatorial structures and their applications, 69–87.
- Boros and Hammer, *Pseudo-Boolean Optimization*, 2002, Discrete Applied Mathematics (123), 155–225.
- Schrijver, *Combinatorial Optimization*, Chapters 44–45.