Linear and quadratic reformulations of nonlinear optimization problems in binary variables

EURO Doctoral Dissertation Award

Elisabeth Rodríguez-Heck

PhD Advisor: Yves Crama

June 25, 2019, Dublin

The problem

Multilinear binary optimization

Set of monomials $\mathcal{S} \subseteq 2^{[n]}$, $a_S \neq 0$ for $S \in \mathcal{S}$.

$$\begin{aligned} \min \sum_{S \in \mathcal{S}} a_S \prod_{i \in S} x_i \\ \text{s. t. } x_i \in \{0, 1\}, \text{ for } i = 1, \ldots, n \end{aligned}$$

Example:

$$\begin{aligned} \min \quad & 9x_1x_2x_3x_4x_5 - 8x_1x_2x_3x_4 - 7x_2x_4x_5 + 9x_1x_3 - 2x_1 - 4x_5 \\ \text{s. t.} \quad & x_1, x_2, x_3, x_4, x_5 \in \{0, 1\} \end{aligned}$$

Linear and quadratic reformulations

Multilinear 0-1 optimization

$\min_{x_i} \sum_{S \in \mathcal{S}} a_S \prod_{i \in S} x_i$

reformulate as

reformulate as

Linear problem

$\min_{x_i, y_j} l(x, y)$

additional y variables

constraints

Quadratic problem

$\min_{x_i, y_j} q(x, y)$

additional y variables

without constraints (choice)

Linear and quadratic reformulations

Multilinear 0-1 optimization

$\min_{x_i} \sum_{S \in \mathcal{S}} a_S \prod_{i \in S} x_i$

reformulate as

reformulate as

Linear problem

$\min_{x_i, y_j} l(x, y)$

additional y variables

constraints

Quadratic problem

$\min_{x_i, y_j} q(x, y)$

additional y variables

without constraints (choice)

The standard linearization (SL)

Nonlinear problem

$$\min \sum_{S \in \mathcal{S}} a_S \prod_{i \in S} x_i + \sum_{i=1}^{n} c_i x_i$$

Linearized problem

$$\min \sum_{S \in \mathcal{S}} a_S y_S + \sum_{i=1}^{n} c_i x_i$$

Standard Linearization (Fortet, 1959; Glover & Woolsey, 1973)

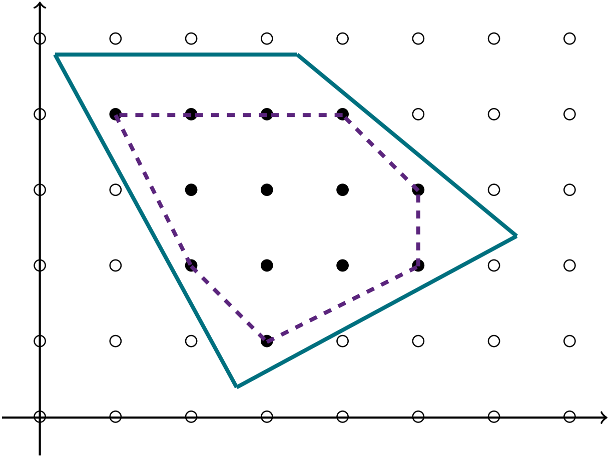
$$y_S = \prod_{i \in S} x_i$$

When $x_i \in \{0, 1\}$ it is equivalent to say:

$$y_S \leq x_i \qquad \forall i \in S, \forall S \in \mathcal{S} \qquad (1)$$

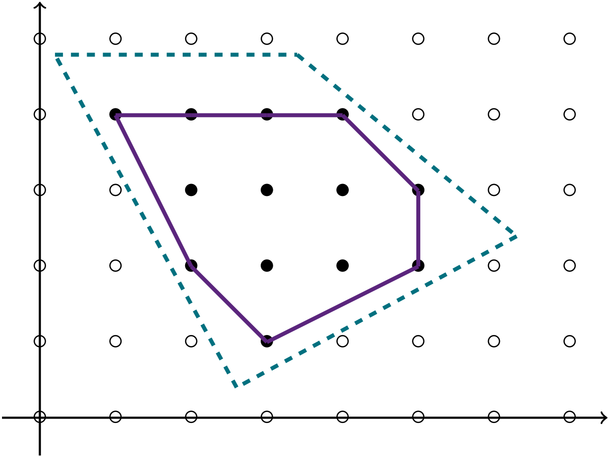
$$y_S \geq \sum_{i \in S} x_i - (|S| - 1) \qquad \forall S \in \mathcal{S} \qquad (2)$$

SL drawback: The continuous relaxation is very weak!



(Image from A. Schrijver's *Theory of linear and integer programming* cover.)

Particular cases for which SL defines the convex hull?



Characterization

Obtained independently, simultaneously: (Del Pia & Khajavirad, 2018).

Theorem 1: (Buchheim, Crama, & Rodríguez-Heck, 2019)

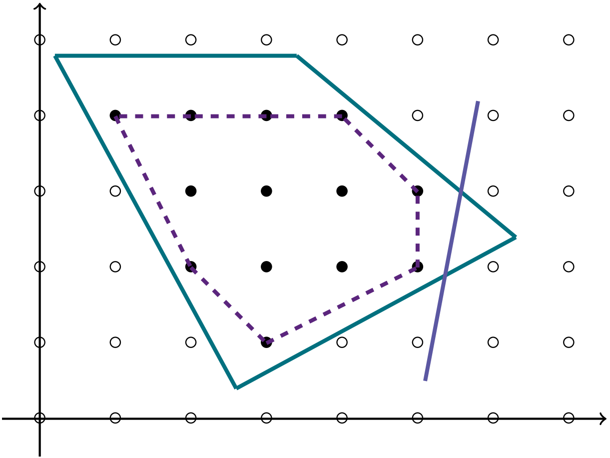
Given a set of monomials $\mathcal{S}$, the following statements are equivalent:

(a) The SL inequalities on $\mathcal{S}$ define an integer polytope.

(b) The matrix of coefficients of the SL inequalities on $\mathcal{S}$ is balanced.

(c) The hypergraph $\mathcal{S}$ is Berge-acyclic.

- The balancedness condition can be checked in polynomial time!



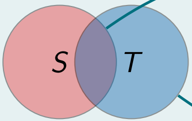
The 2-link inequalities

Definition (Crama & Rodríguez-Heck, 2017)

For $S, T \in \mathcal{S}$ and y_S, y_T such that $y_S = \prod_{i \in S} x_i$, $y_T = \prod_{i \in T} x_i$, the **2-link** associated with (S, T) is the linear inequality

$$\boxed{\mathbf{y_S \leq y_T - \sum_{i \in T \setminus S} x_i + |T \setminus S|}}$$

Interpretation



$$y_S = 1 \Rightarrow \forall i \in S, x_i = 1$$

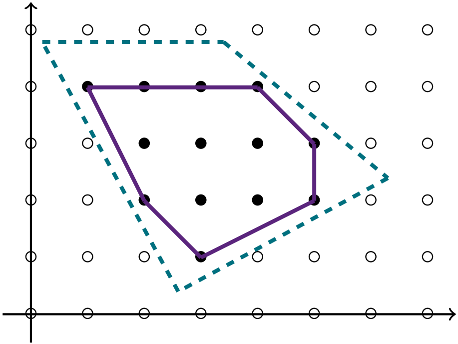
$$y_T = 0 \text{ and } y_S = 1 \Rightarrow \exists j \in T \setminus S, x_j = 0$$

- The 2-links are only in quadratic number in $|\mathcal{S}|$.

A complete description for the case of two monomials

Theorem 2: (Crama & Rodríguez-Heck, 2017)

For the case of two nonlinear monomials the **standard linearization** and the **2-links** provide a complete description of the **convex hull**.



Linear and quadratic reformulations

Multilinear 0-1 optimization

$\min_{x_i} \sum_{S \in \mathcal{S}} a_S \prod_{i \in S} x_i$

reformulate as

reformulate as

Linear problem

$\min_{x_i, y_j} l(x, y)$

additional y variables

constraints

Quadratic problem

$\min_{x_i, y_j} q(x, y)$

additional y variables

without constraints (choice)

Quadratization: definition and desirable properties

Definition (Anthony, Boros, Crama, & Gruber, 2017)

Given a multilinear polynomial $f(x)$ where $x \in \{0,1\}^n$, a *quadratization* $g(x,y)$ is a function satisfying

- g is quadratic
- $g(x,y)$ depends on the original variables x and on m auxiliary variables y
- satisfies

$$f(x) = \min\{g(x,y) : y \in \{0,1\}^m\} \quad \forall x \in \{0,1\}^n.$$

Which quadratizations are "good"?

- Small number of auxiliary variables (*compact*).
- Small number of positive quadratic terms ($x_i x_j$, $x_i y_j$ ...) (empirical distance from submodularity).
- ...

Termwise quadratizations

Termwise quadratizations

Main idea

Quadratize monomial by monomial using disjoint sets of auxiliary variables.

$$f(x) = -35x_1x_2x_3x_4x_5 + 50x_1x_2x_3x_4 - 10x_1x_2x_4x_5 + 5x_2x_3x_4 + 5x_4x_5 - 20x_1$$

Negative monomial

(Kolmogorov & Zabih, 2004; Freedman & Drineas, 2005)

$$-\prod_{i=1}^{n} x_i = \min_{y \in \{0,1\}} -y(\sum_{i=1}^{n} x_i - (n-1))$$

- One variable is sufficient.
- No positive quadratic terms.

Check that, for every $x \in \{0,1\}^n$, $min_y g(x,y) = -\prod_{i=1}^{n} x_i$., two cases:

1. If $x_i = 1\ \forall i$, then $min_y - y$, minimum value of -1 reached for $y = 1$.
2. If $\exists i$ such that $x_i = 0$, then $min_y - Cy$, where $C \leq 0$, minimum value of 0 reached for $y = 0$.

Termwise quadratizations

Main idea

Quadratize monomial by monomial using disjoint sets of auxiliary variables.

$$f(x) = -35x_1x_2x_3x_4x_5 + 50x_1x_2x_3x_4 - 10x_1x_2x_4x_5 + 5x_2x_3x_4 + 5x_4x_5 - 20x_1$$

Negative monomial

(Kolmogorov & Zabih, 2004; Freedman & Drineas, 2005)

$$-\prod_{i=1}^{n} x_i = \min_{y\in\{0,1\}} -y\left(\sum_{i=1}^{n} x_i - (n-1)\right)$$

- One variable is sufficient.
- No positive quadratic terms.

Positive monomial

(Ishikawa, 2011)

$$\prod_{i=1}^{n} x_i = \min_{y\in\{0,1\}^k} \sum_{i=1}^{k} y_i(c_{i,n}(-|x| + 2i) - 1) + \frac{|x|(|x|-1)}{2},$$

- Number of variables: $\mathbf{k} = \left\lfloor \frac{\mathbf{n-1}}{\mathbf{2}} \right\rfloor$.
- $\binom{n}{2}$ positive quadratic terms.

Upper bound for the positive monomial: $\lceil \log(n) \rceil - 1$

Theorem 3 (Boros, Crama, & Rodríguez-Heck, 2018)

Assume that $n = 2^\ell$ and let $|x| = \sum_{i=1}^{n} x_i$ be the Hamming weight of $x \in \{0,1\}^n$. Then,

$$g(x,y) = \frac{1}{2}(|x| - \sum_{i=1}^{\ell-1} 2^i y_i)(|x| - \sum_{i=1}^{\ell-1} 2^i y_i - 1)$$

is a quadratization of the positive monomial $P_n(x) = \prod_{i=1}^{n} x_i$ using $\lceil \log(\mathbf{n}) \rceil - \mathbf{1}$ auxiliary variables.

Where does $\lceil \log(\mathbf{n}) \rceil - \mathbf{1}$ come from?

- The quadratization depends on $|x|$, which takes values between 0 and n.
- Integers between 0 and $n-1$ can be represented as a sum of $\log(n)$ powers of 2.
- Use y variables to express which powers of 2 are in the sum.
- Use one factor to represent *odd* $|x|$ and the other factor for *even* $|x|$.

Lower bound for the positive monomial

Theorem 4 (Boros, Crama, & Rodríguez-Heck, 2018)

If $g(x, y)$ is a quadratization of the positive monomial $P_n(x) = \prod_{i=1}^{n} x_i$ using m variables, then

$$m \geq \lceil \log(n) \rceil - 1$$

Pairwise covers

Pairwise covers

Anthony, Boros, Crama and Gruber (2017)

Substituting common sets of variables

$$f(x) = -35x_1x_2x_3x_4x_5 + 50x_1x_2x_3x_4 - 10x_1x_2x_4x_5 + 5x_2x_3x_4 + 5x_4x_5 - 20x_1$$

could be replaced by

$$f(x) = -35y_{12}y_{345} + 50y_{12}y_{34} - 10y_{12}y_{45} + 5x_2y_{34} + 5x_4x_5 - 20x_1 + P(x, y)$$

where $P(x, y)$ imposes $y_{12} = x_1x_2$, $y_{345} = y_{34}x_5$...

- Defining Pairwise Covers with smallest number of y variables is NP-hard.
- Three heuristics: PC1, PC2, PC3, based on the idea of substituting subsets of variables that appear *more frequently* in the monomial set $\mathcal{S}$ *with higher priority*.

Linear and quadratic reformulations

Multilinear 0-1 optimization

$\min_{x_i} \sum_{S \in \mathcal{S}} a_S \prod_{i \in S} x_i$

reformulate as | reformulate as

Linear problem

$\min_{x_i, y_j} l(x, y)$

additional y variables

constraints

Quadratic problem

$\min_{x_i, y_j} q(x, y)$

additional y variables

without constraints (choice)

Comparing linear and quadratic reformulations computationally

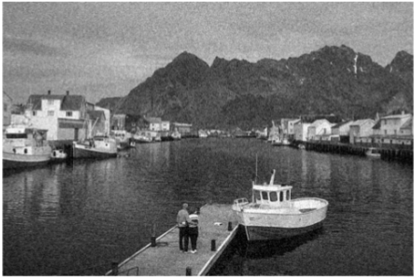
Methods and Scope

Methods						
Linearizations		Quadratizations				
	2-links	*Pairw. cov.*			*Termwise*	
SL	SL-2L	PC1	PC2	PC3	Ishikawa	logn-1

- We use CPLEX 12.7 as solver for reformulated linear and quadratic problems.
- Is this the best choice? Topic for discussion...
- Next steps: test other solvers.

Application in computer vision: image restoration

Input: blurred image



Output: restored image

Image from the Corel database.

Instances: Vision

Image restoration

1	0	0	0	0	0
0	0	1	1	0	0
0	1	1	0	1	0
0	1	1	1	1	0
0	0	1	1	0	1
0	0	0	0	0	0

$\longrightarrow$

0	0	0	0	0	0
0	0	1	1	0	0
0	1	1	1	1	0
0	1	1	1	1	0
0	0	1	1	0	0
0	0	0	0	0	0

- Variables: output pixel
- Minimize energy
- Nonlinear: penalize "non-natural" configurations

$\begin{matrix} 1 & 1 \\ 1 & 1 \end{matrix}$ term: $10x_1x_2x_3x_4$

$\begin{matrix} 0 & 1 \\ 1 & 0 \end{matrix}$ term: $40(1-x_1)x_2x_3(1-x_4)$

$\begin{matrix} 0 & 1 \\ 0 & 0 \end{matrix}$ term: $20(1-x_1)x_2(1-x_3)(1-x_4)$

Instances: Vision

Image restoration

$$
\begin{matrix}
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 \\
0 & 1 & 1 & 0 & 1 & 0 \\
0 & 1 & 1 & 1 & 1 & 0 \\
0 & 0 & 1 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0
\end{matrix}
\longrightarrow
\begin{matrix}
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 \\
0 & 1 & 1 & 1 & 1 & 0 \\
0 & 1 & 1 & 1 & 1 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0
\end{matrix}
$$

Base images:

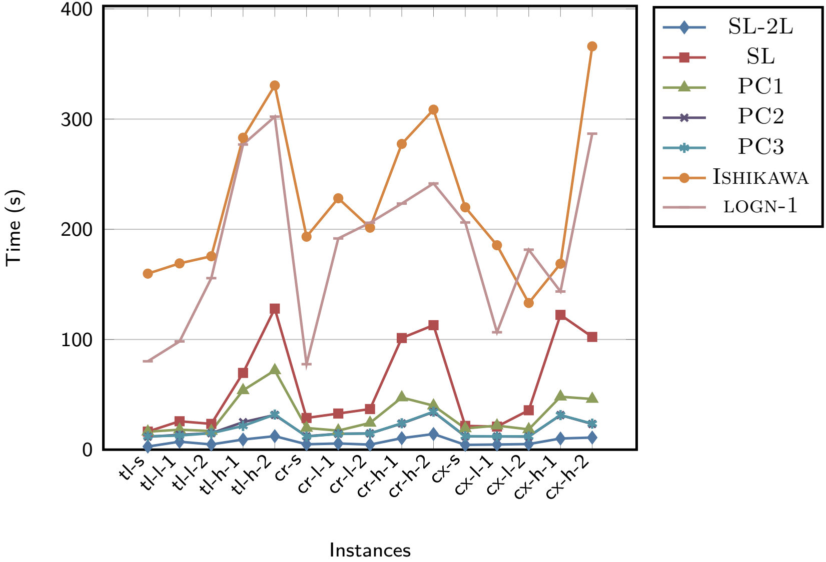
- top left rect. (tl)
- centre rect. (cr)
- cross (cx)

Perturbations:

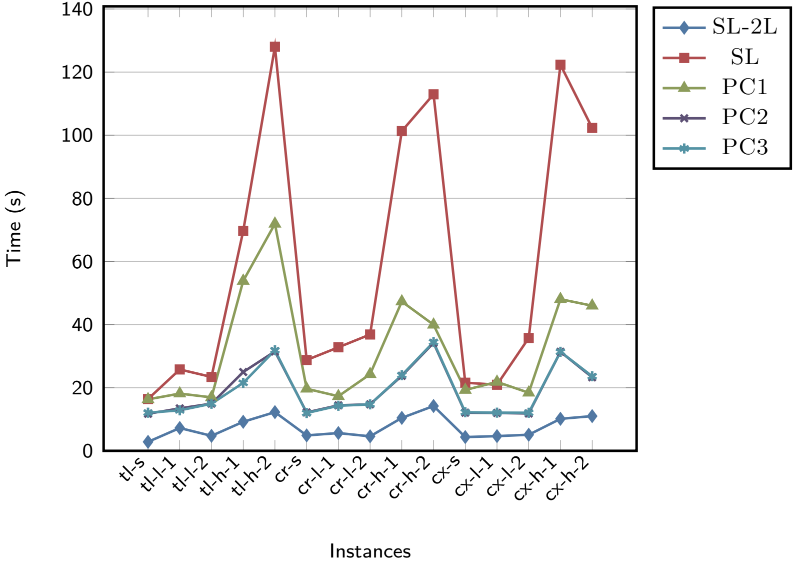
- none (n)
- low (l)
- high (h)

Up to $n = 900$ variables and $m = 6788$ terms

Vision: all methods 15 × 15 ($n = 225$, $m = 1598$)



Vision: best methods 15 × 15 ($n = 225$, $m = 1598$)



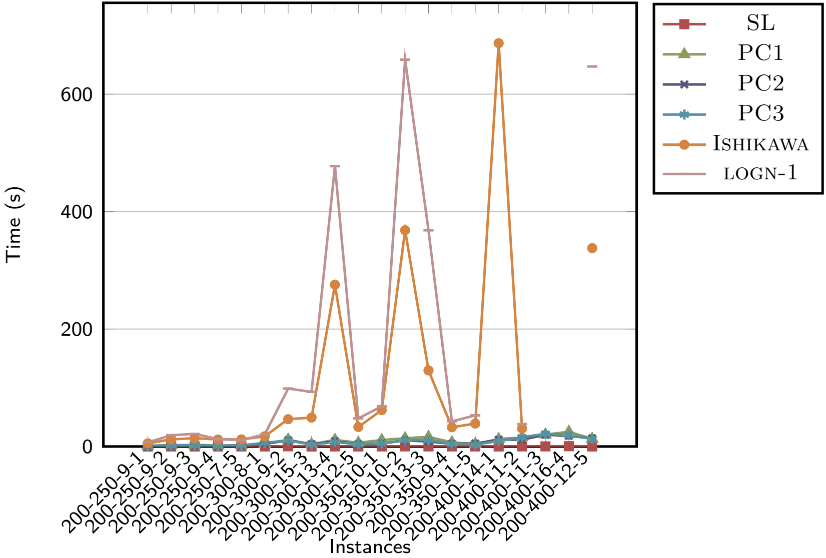
Vision: Quadratization properties

	Pairwise covers	Termwise
Number of y variables	less	more
Number of positive quadratic terms	less	more

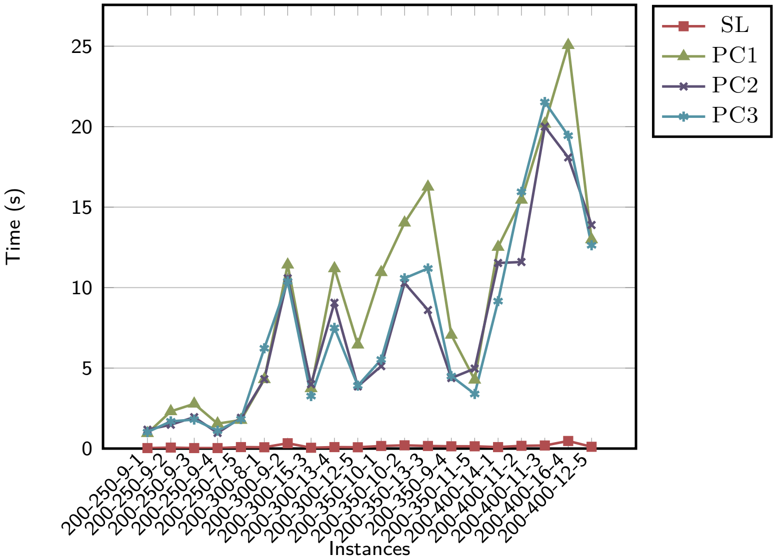
Instances

- Similar behavior observed for a different set of instances with special structure.
- What happens with random polynomials?
- Generated random polynomials of degrees between 7 and 17 (as in (Buchheim & Rinaldi, 2007)).

Random high degree: all methods



Random high degree: best methods



Random high degree: Quadratization properties

	Pairwise covers	Termwise
Number of y variables	**more**	**less**
Number of positive quadratic terms	less	more

Summary of results

- Suprising behaviour for vision instances:
  - Pairwise covers faster than SL default CPLEX branch & cut.
  - SL with 2-links faster than pairwise covers.
- Termwise quadratizations are consistently slower than pairwise covers...
- ...even when they use smaller number of y variables.

Some perspectives

Theoretical:

1. Understand better which properties define a "good" quadratization.
2. Which formulation is better: SL or SL of a quadratization?

Experimental:

1. Experiments on further applications.
2. Currently working on open-sourcing reformulations code.
3. Repeat experiments with other solvers (convexification, SDP...).

... and many others!

Thank you for your attention!

Elisabeth Rodríguez-Heck

rodriguez-heck@or.rwth-aachen.de

Anthony, M., Boros, E., Crama, Y., & Gruber, A. (2017). Quadratic reformulations of nonlinear binary optimization problems. *Mathematical Programming*, *162*(1-2), 115–144.

Boros, E., Crama, Y., & Rodríguez-Heck, E. (2018). *Compact quadratizations for pseudo-Boolean functions.* (Submitted)

Buchheim, C., Crama, Y., & Rodríguez-Heck, E. (2019). Berge-acyclic multilinear 0-1 optimization problems. *European Journal of Operational Research*, *273*, 102–107.

Buchheim, C., & Rinaldi, G. (2007). Efficient reduction of polynomial zero-one optimization to the quadratic case. *SIAM Journal on Optimization*, *18*(4), 1398–1413.

Crama, Y., & Rodríguez-Heck, E. (2017). A class of valid inequalities for multilinear 0-1 optimization problems. *Discrete Optimization*, *25*, 28–47.

Bibliography II

Del Pia, A., & Khajavirad, A. (2018). The multilinear polytope for acyclic hypergraphs. *SIAM Journal on Optimization*, *28*(2), 1049–1076.

Fortet, R. (1959). L'algèbre de Boole et ses applications en recherche opérationnelle. *Cahiers du Centre d'Études de Recherche Opérationnelle*, *4*, 5–36.

Freedman, D., & Drineas, P. (2005, June). Energy minimization via graph cuts: settling what is possible. In *Ieee conference on computer vision and pattern recognition* (Vol. 2, pp. 939–946).

Glover, F., & Woolsey, E. (1973). Further reduction of zero-one polynomial programming problems to zero-one linear programming problems. *Operations Research*, *21*(1), 156–161.

Bibliography III

Ishikawa, H. (2011, June). Transformation of general binary MRF minimization to the first-order case. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *33*(6), 1234–1249.

Kolmogorov, V., & Zabih, R. (2004, Feb). What energy functions can be minimized via graph cuts? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *26*(2), 147–159.