Linear and quadratic reformulations of nonlinear optimization problems in binary variables

Elisabeth Rodríguez-Heck
RWTH Aachen University, Chair of Operations Research

Results of PhD thesis
advised by Yves Crama, University of Liège

G-Scop, Grenoble, February 7, 2019

Introduction

Pseudo-Boolean optimization

General problem: pseudo-Boolean optimization

Given a pseudo-Boolean function $f : \{0,1\}^n \to \mathbb{R}$

$$\min_{x \in \{0,1\}^n} f(x).$$

Pseudo-Boolean optimization

General problem: pseudo-Boolean optimization

Given a pseudo-Boolean function $f : \{0,1\}^n \to \mathbb{R}$

$$\min_{x \in \{0,1\}^n} f(x).$$

Theorem (Hammer, Rosenberg, & Rudeanu, 1963)

Every pseudo-Boolean function $f : \{0,1\}^n \to \mathbb{R}$ admits a unique multilinear expression.

Pseudo-Boolean optimization

General problem: pseudo-Boolean optimization

Given a pseudo-Boolean function $f : \{0,1\}^n \to \mathbb{R}$

$$\min_{x \in \{0,1\}^n} f(x).$$

Theorem (Hammer, Rosenberg, & Rudeanu, 1963)

Every pseudo-Boolean function $f : \{0,1\}^n \to \mathbb{R}$ admits a unique multilinear expression.

- Given f, finding its unique multilinear representation can be costly! (Size of the input: $O(2^n)$)

Multilinear 0–1 optimization

Assumption: f given as a multilinear polynomial

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

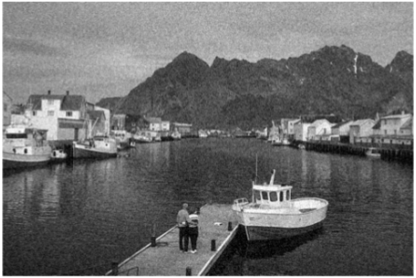
$$
\begin{aligned}
\min \sum_{S \in \mathcal{S}} a_S \prod_{i \in S} x_i \\
\text{s. t. } x_i \in \{0, 1\}, \text{ for } i = 1, \ldots, n
\end{aligned}
$$

Example:

$$f(x_1, x_2, x_3) = 9x_1x_2x_3 + 8x_1x_2 - 6x_2x_3 - 2x_1$$

Application in computer vision: image restoration

Input: blurred image



Output: restored image

Image from the Corel database.

Applications

- Location problems
- Joint supply chain design and inventory management
- Statistical mechanics
- Quantum computing
- ...

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

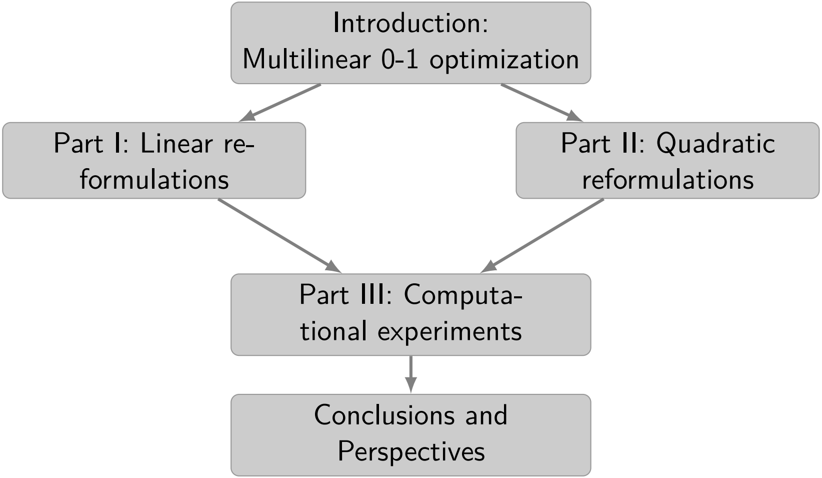
Quadratic problem

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

additional y variables

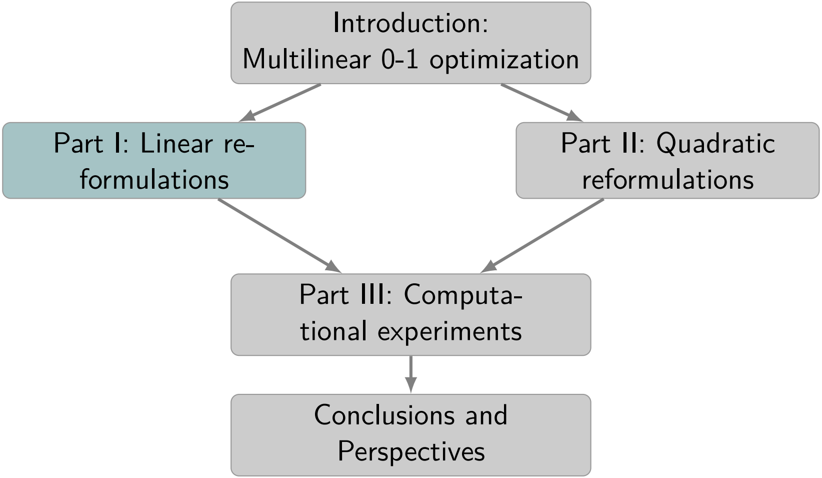
without constraints (choice)

Overview



Introduction:
Multilinear 0-1 optimization
Part I: Linear reformulations
Part II: Quadratic reformulations
Part III: Computational experiments
Conclusions and Perspectives

Overview


Introduction:
Multilinear 0-1 optimization
Part I: Linear re-
formulations
Part II: Quadratic
reformulations
Part III: Computa-
tional experiments
Conclusions and
Perspectives

Introduction to linearizations

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)

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

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:

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

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

SL drawback: The continuous relaxation given by the SL is very weak!

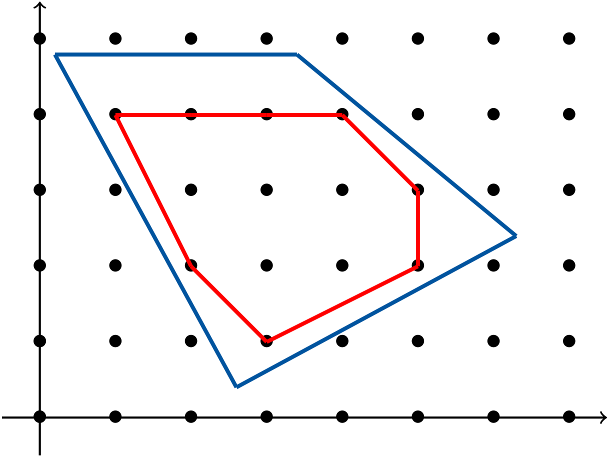
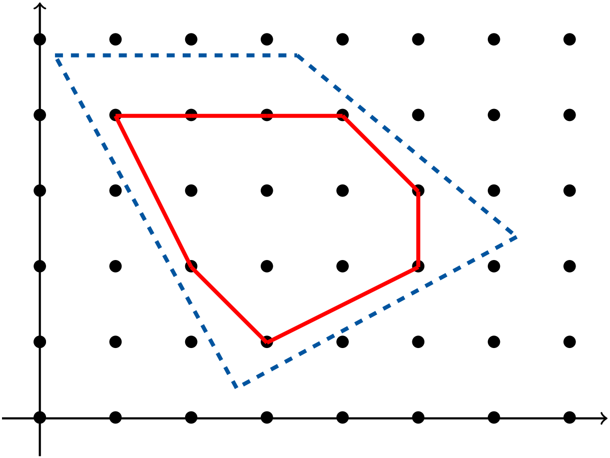


Image from A. Schrijver's book cover:
Theory of linear and integer programming.

When do the SL inequalities provide a perfect formulation?



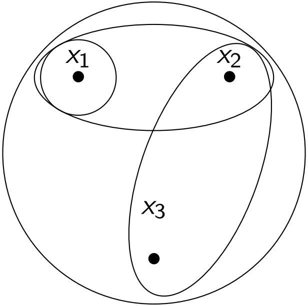
Hypergraph associated with a polynomial

A multilinear polynomial can be associated to a hypergraph $H = (V = \{1, \dots, n\}, \mathcal{S})$.

$$\min \sum_{S \in \mathcal{S}} a_S \prod_{i \in S} x_i + \sum_{i \in V} c_i x_i$$

Example:

$$f(x_1, x_2, x_3) = 9x_1x_2x_3 - 8x_1x_2 + 6x_2x_3 - 2x_1$$



Characterization for the unsigned case

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

Given a hypergraph H, the following statements are equivalent:

(a) The SL inequalities define an integer polytope.

(b) The matrix of coefficients of the SL inequalities is balanced.

(c) The hypergraph H is Berge-acyclic.

Characterization for the unsigned case

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

Given a hypergraph H, the following statements are equivalent:

(a) The SL inequalities define an integer polytope.

(b) The matrix of coefficients of the SL inequalities is balanced.

(c) The hypergraph H is Berge-acyclic.

- Obtained independently and simultaneously by (Del Pia & Khajavirad, 2018).

Characterization for the unsigned case

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

Given a hypergraph H, the following statements are equivalent:

(a) The SL inequalities define an integer polytope.

(b) The matrix of coefficients of the SL inequalities is balanced.

(c) The hypergraph H is Berge-acyclic.

- Obtained independently and simultaneously by (Del Pia & Khajavirad, 2018).
- Theorem 1 is a Corollary of a more general result considering the signs of the coefficients.

Characterization for the unsigned case

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

Given a hypergraph H, the following statements are equivalent:

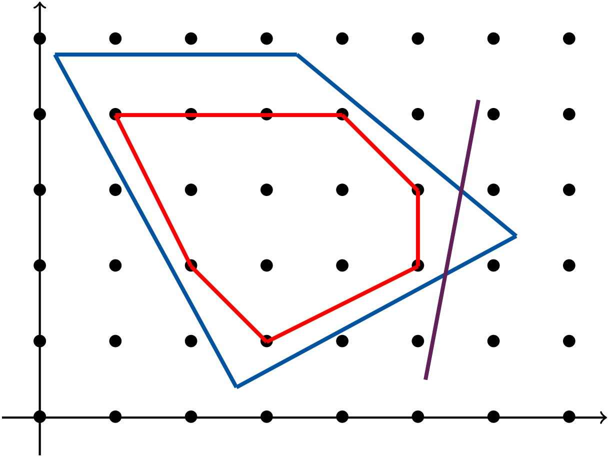
(a) The SL inequalities define an integer polytope.

(b) The matrix of coefficients of the SL inequalities is balanced.

(c) The hypergraph H is Berge-acyclic.

- Obtained independently and simultaneously by (Del Pia & Khajavirad, 2018).
- Theorem 1 is a Corollary of a more general result considering the signs of the coefficients.
- The balancedness condition can be checked in polynomial time.

A class of valid inequalities for multilinear 0–1 optimization problems



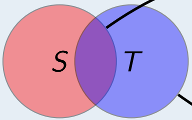
The 2-link inequalities

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



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

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

A complete description for the case of two monomials

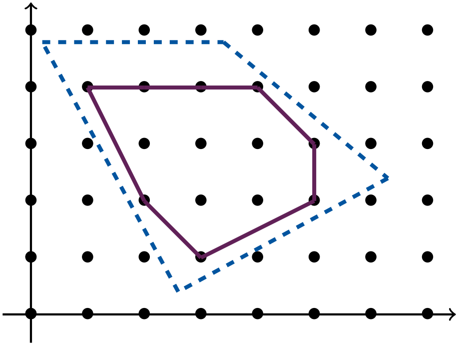
Theorem 2: (Crama & Rodríguez-Heck, 2017)

For the case of two nonlinear monomials, $P^*_{SL} = P^{2links}_{SL}$, i.e., the **standard linearization** and the **2-links** provide a **complete description** of P^*_{SL}.

A complete description for the case of two monomials

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

For the case of two nonlinear monomials, $P_{SL}^* = P_{SL}^{2links}$, i.e., the **standard linearization** and the **2-links** provide a **complete description** of P_{SL}^*.



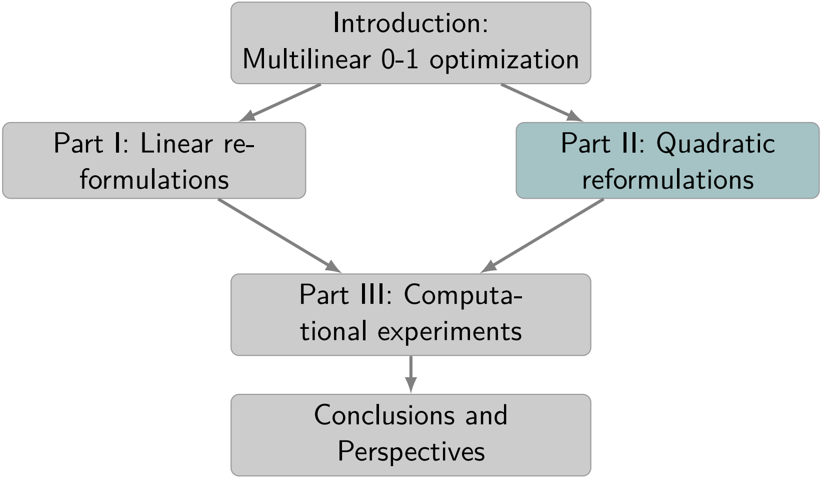
Proof idea: $y_{S\cap T} = \prod_{i\in S\cap T} x_i$, $y_S = y_{S\cap T}\prod_{i\in S\setminus T} x_i$, $y_T = y_{S\cap T}\prod_{i\in T\setminus S} x_i$

- Consider the extended formulation (with variables in $[0,1]$)

$$
\begin{aligned}
y_{S\cap T} &\le x_i, && \forall i \in S\cap T, && (3)\\
y_{S\cap T} &\ge \sum_{i\in S\cap T} x_i - (|S\cap T| - 1), && && (4)\\
y_S &\le y_{S\cap T}, && && (5)\\
y_S &\le x_i, && \forall i \in S\setminus T, && (6)\\
y_S &\ge \sum_{i\in S\setminus T} x_i + y_{S\cap T} - |S\setminus T|, && && (7)\\
y_T &\le y_{S\cap T}, && && (8)\\
y_T &\le x_i, && \forall i \in T\setminus S, && (9)\\
y_T &\ge \sum_{i\in T\setminus S} x_i + y_{S\cap T} - |T\setminus S|, && && (10)
\end{aligned}
$$

- Notice that the two polytopes P^0 and P^1 obtained by fixing variable $y_{S\cap T}$ to 0 and 1, resp., are integral.
- Compute $conv(P^0 \cup P^1)$ using (Balas, 1974) and see that it is P_{SL}^{2links}.

Overview


Introduction:
Multilinear 0-1 optimization
Part I: Linear reformulations
Part II: Quadratic reformulations
Part III: Computational experiments
Conclusions and Perspectives

Introduction to quadratizations

Quadratization: definition and desirable properties

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

Given a pseudo-Boolean function $f(x)$ 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

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

Quadratization: definition and desirable properties

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

Given a pseudo-Boolean function $f(x)$ 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"?

Quadratization: definition and desirable properties

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

Given a pseudo-Boolean function $f(x)$ 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*).

Quadratization: definition and desirable properties

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

Given a pseudo-Boolean function $f(x)$ 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).

Quadratization: definition and desirable properties

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

Given a pseudo-Boolean function $f(x)$ 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).
- Set of quadratic terms with specific underlying graphs.

Quadratization: definition and desirable properties

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

Given a pseudo-Boolean function $f(x)$ 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).
- Set of quadratic terms with specific underlying graphs.
- ...

Persistencies

Weak Persistency Theorem (Hammer, Hansen, & Simeone, 1984)

Let (QP) be a quadratic optimization problem on $x \in \{0,1\}^n$, and let $(\tilde{x}, \tilde{y})$ be an optimal solution of the *continuous standard linearization of* (QP)

$$
\begin{aligned}
\min \quad & c_0 + \sum_{j=1}^{n} c_j x_j + \sum_{1 \le i < j \le n} c_{ij} y_{ij} & \\
\text{s. t.} \quad & y_{ij} \ge x_i + x_j - 1 & i,j = 1,\dots,n, i < j \\
& y_{ij} \le x_i & i,j = 1,\dots,n, i < j \\
& y_{ij} \le x_j & i,j = 1,\dots,n, i < j \\
& 0 \le y_{ij} \le 1 & i,j = 1,\dots,n, i < j \\
& 0 \le x_i \le 1 & i = 1,\dots,n
\end{aligned}
$$

such that $\tilde{x}_j = 1$ for $j \in O$ and $\tilde{x}_j = 0$ for $j \in Z$. Then, for any minimizing vector x^* of (QP) switching $x_j^* = 1$ for $j \in O$ and $x_j^* = 0$ for $j \in Z$ will also yield a minimum of f.

(See also survey (Boros & Hammer, 2002).)

Persistencies

- The Weak Persistency Theorem is not the strongest form of persistency.

Persistencies

- The Weak Persistency Theorem is not the strongest form of persistency.
- There are ways to compute, in polynomial time, a *maximal* set of variables to fix, based on a network flow algorithm (Boros, Hammer, Sun, & Tavares, 2008).

Persistencies

- The Weak Persistency Theorem is not the strongest form of persistency.
- There are ways to compute, in polynomial time, a *maximal* set of variables to fix, based on a network flow algorithm (Boros, Hammer, Sun, & Tavares, 2008).
- In computer vision, image restoration and related problems of up to *millions* of variables are efficiently solved, thanks to the use of persistencies.

Termwise quadratizations

Termwise quadratizations

Main idea

Quadratize monomial by monomial using disjoint sets of auxiliary variables.

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

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.

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:

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 $y = 1$.

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: $k = \lfloor \frac{n-1}{2} \rfloor$.
- $\binom{n}{2}$ positive quadratic terms.

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

Theorem 3 (simplified) (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 $\lceil \log(n) \rceil - 1$ auxiliary variables.

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

Theorem 3 (simplified) (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 $\lceil \log(n) \rceil - 1$ auxiliary variables.

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

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

Theorem 3 (simplified) (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 $\lceil \log(n) \rceil - 1$ auxiliary variables.

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

- The quadratization depends on $|x|$, which takes values between 0 and n.

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

Theorem 3 (simplified) (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 $\lceil \log(n) \rceil - 1$ auxiliary variables.

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

- The quadratization depends on $|x|$, which takes values between 0 and n.
- Case 1 ($|x| \leq n-1$): Integers between 0 and $n-1$ can be represented as a sum of $\log(n)$ powers of 2.

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

Theorem 3 (simplified) (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 $\lceil \log(n) \rceil - 1$ auxiliary variables.

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

- The quadratization depends on $|x|$, which takes values between 0 and n.
- Case 1 ($|x| \leq n-1$): 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.

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

Theorem 3 (simplified) (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 $\lceil \log(n) \rceil - 1$ auxiliary variables.

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

- The quadratization depends on $|x|$, which takes values between 0 and n.
- Case 1 ($|x| \leq n-1$): 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.
- For $|x| \leq n-1$, one factor to reach the minimum value of zero for *odd* $|x|$ and the other factor for *even* $|x|$.

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

Theorem 3 (simplified) (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 $\lceil \log(n) \rceil - 1$ auxiliary variables.

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

- The quadratization depends on $|x|$, which takes values between 0 and n.
- Case 1 ($|x| \leq n-1$): 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.
- For $|x| \leq n-1$, one factor to reach the minimum value of zero for *odd* $|x|$ and the other factor for *even* $|x|$.
- Case 2 ($|x| = n$): Similarly, we can show $min_y g(x,y) = 1$.

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

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

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

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

Proof idea:

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

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

Proof idea:

- Consider $r(x) = \prod_{y \in \{0,1\}^m} g(x, y)$.

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

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

Proof idea:

- Consider $r(x) = \prod_{y \in \{0,1\}^m} g(x, y)$.
- $\deg(r) \leq 2 \cdot 2^m$.

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

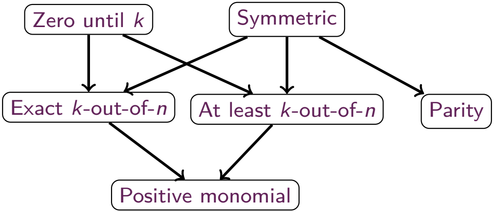
$$m \geq \lceil \log(n) \rceil - 1$$

Proof idea:

- Consider $r(x) = \prod_{y \in \{0,1\}^m} g(x, y)$.
- $\deg(r) \leq 2 \cdot 2^m$.
- $\deg(r) \geq n$, because $r(x) = \alpha P_n(x)$ where $\alpha > 0$ (unicity of the multilinear representation). More precisely,
  - If $|x| < n$, there exists $y \in \{0,1\}^m$ such that $g(x, y) = 0$.
  - If $|x| = n$, $g(x, y) \geq 1$ for all $y \in \{0,1\}^m$.

Results for more general functions

Function	Lower Bound	Upper Bound
Zero until k	$\Omega(2^{\frac{n}{2}})$ for some function[1] $\lceil \log(k) \rceil - 1$ for all functions	$O(2^{\frac{n}{2}})$ [1]
Symmetric	$\Omega(\sqrt{n})$ for some function[2]	$O(\sqrt{n}) = 2\lceil \sqrt{n+1} \rceil$
Exact k-out-of-n	$\max(\lceil \log(k) \rceil, \lceil \log(n-k) \rceil) - 1$	$\max(\lceil \log(k) \rceil, \lceil \log(n-k) \rceil)$
At least k-out-of-n	$\lceil \log(k) \rceil - 1$	$\max(\lceil \log(k) \rceil, \lceil \log(n-k) \rceil)$
Positive monomial	$\lceil \log(n) \rceil - 1$	$\lceil \log(n) \rceil - 1$
Parity	$\lceil \log(n) \rceil - 1$	$\lceil \log(n) \rceil - 1$



[1] see (Anthony et al., 2017)
[2] see (Anthony, Boros, Crama, & Gruber, 2016)

RWTH AACHEN UNIVERSITY

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 $y_{12} = x_1x_2$, $y_{345} = y_{34}x_5$...

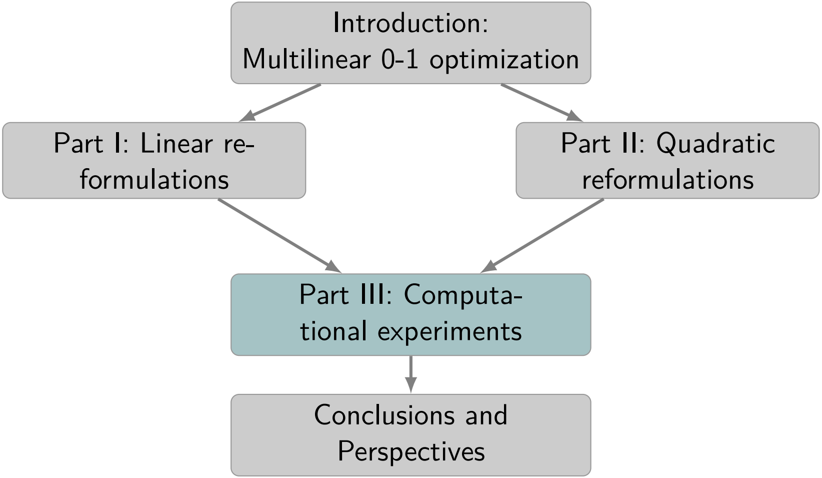
Heuristics for small Pairwise Covers

Three heuristics:

- PC1: Separate first two variables from the rest.
- PC2: Most "popular" intersections first.
- PC3: Most "popular" pairs of variables first.

Main idea: identifying *subterms* that appear as subsets of one or more monomials in the input monomial set $\mathcal{S}$.

Overview


Introduction:
Multilinear 0-1 optimization
Part I: Linear re-formulations
Part II: Quadratic reformulations
Part III: Computa-tional experiments
Conclusions and Perspectives

Comparing linear and quadratic reformulations

Methods

<table>
<tr><th colspan="7">Methods</th></tr>
<tr><td colspan="2">Linearizations</td><td colspan="5">Quadratizations</td></tr>
<tr><td></td><td>2-links</td><td colspan="3">Pairw. cov.</td><td colspan="2">Termwise</td></tr>
<tr><td>SL</td><td>SL-2L</td><td>PC1</td><td>PC2</td><td>PC3</td><td>Ishikawa</td><td>logn-1</td></tr>
</table>

Objectives and scope

Objectives

1. Compare resolution times of linear and quadratic reformulations *when relying on a commercial solver*.
2. Are the results consistent for different types of instances?

The objective is *not* to compete with existing literature.

Objectives and scope

Objectives

1. Compare resolution times of linear and quadratic reformulations *when relying on a commercial solver*.
2. Are the results consistent for different types of instances?

The objective is *not* to compete with existing literature.

Dependence on the underlying solver: CPLEX 12.7

Drawbacks:

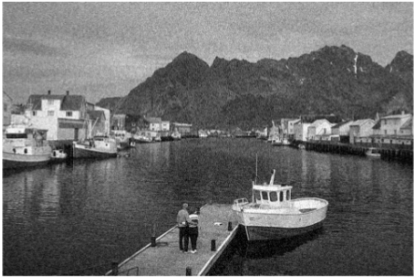
- ✗ Linear solver possibly more powerful than quadratic solver.
- ✗ "Blackbox": little control over the actual resolution.
- ✗ Cannot exploit interesting properties like persistencies.

Advantages:

- ✓ Commercial widely used solver.
- ✓ Avoid comparing commercial against "home-made" software.

Application in computer vision: image restoration

Input: blurred image



Output: restored image

Image from the Corel database.

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:

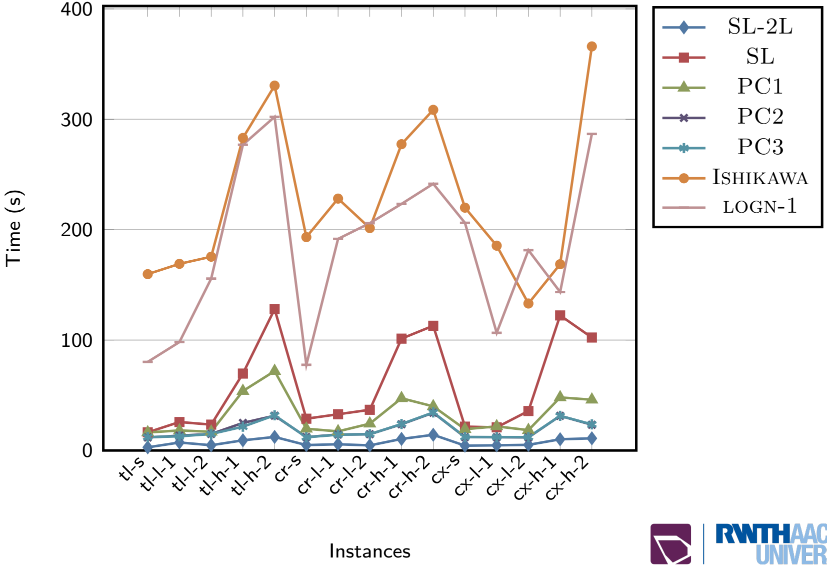
- top left rect. (tl)
- centre rect. (cr)
- cross (cx)

Perturbations:

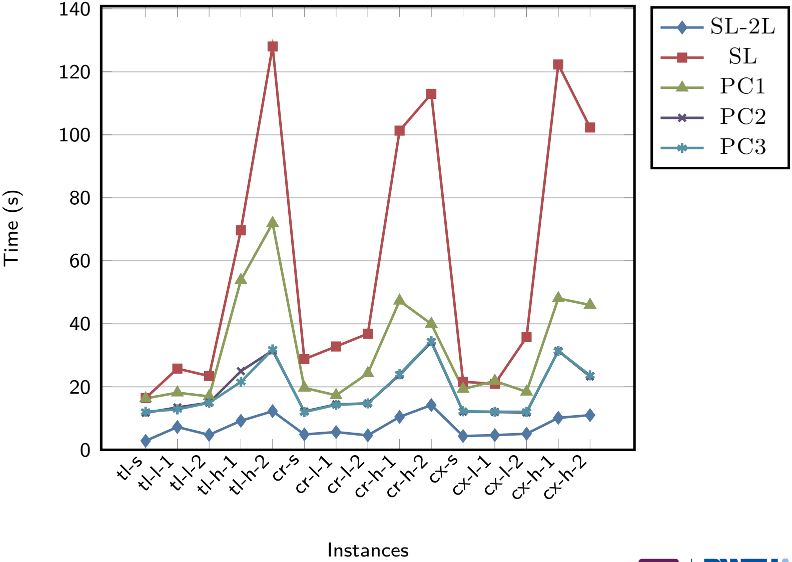
- none (n)
- low (l)
- high (h)

$\boxed{\text{Up to } n = 900 \text{ variables and } m = 6788 \text{ terms}}$

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



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



RWTH AACHEN UNIVERSITY

Vision: Quadratization properties

	Pairwise covers	Termwise
Number of y variables	less	more
Number of positive quadratic terms	less	more
Monomial interactions	✓	✗

Dynamical glass transition phenomenon

Dynamical glass transition phenomenon

Freeze for a few hours
(-12 ° C)

Dynamical glass transition phenomenon

Freeze for a few hours
(-12 ° C)

Dynamical glass transition phenomenon

Freeze for a few hours
(-12 ° C)

Dynamical glass transition phenomenon

Freeze for a few hours
(-12 ° C)

Dynamical glass transition phenomenon

Freeze for a few hours
(-12 ° C)

More on YouTube: supercooled water, Lenape High School - science skills class.
Scientific articles: (Bernasconi, 1987; Liers, Marinari, Pagacz, Ricci-Tersenghi, & Schmitz, 2010).
Instances downloaded from
`http://polip.zib.de/autocorrelated_sequences/`.

Autocorrelated sequences

At least one method finished in 1h for instances in **bold**: 8/40 (!).

Id	n	m	Id	n	m
20_5	**20**	**207**	45_5	45	507
20_10	**20**	**833**	45_11	45	2 813
25_6	**25**	**407**	45_23	45	10 776
25_13	**25**	**1 782**	45_34	45	18 348
25_19	25	3 040	45_45	45	21 993
25_25	25	3 677	50_6	50	882
30_4	**30**	**223**	50_13	50	4 457
30_8	**30**	**926**	50_25	50	14 412
30_15	30	2 944	50_38	50	25 446
30_23	30	5 376	50_50	50	30 271
30_30	30	6 412	55_6	55	977
35_4	**35**	**263**	55_14	55	5 790
35_9	35	1 381	55_28	55	19 897
35_18	35	5 002	55_41	55	33 318
35_26	35	8 347	55_55	55	40 402
40_5	**40**	**447**	60_8	60	2 036
40_10	40	2 053	60_15	60	7 294
40_20	40	7 243	60_30	60	25 230
40_30	40	12 690	60_45	60	43 689
40_40	40	15 384	60_60	60	52 575

Autocorrelated sequences: preliminary results

Instance			Resolution time (secs)			
Id	n	m	SL-2L	SL	SL-2L-ONLY	SL-NOCUTS
20_5	20	207	6.94	11.48	**1.06**	6.13
20_10	20	833	112.27	91.75	**47**	65.89
25_6	25	407	65.36	137.38	**44.52**	321.77
30_4	30	223	**4.24**	15.7	13.47	349.84
35_4	35	263	**11.92**	36.86	69.03	3104.45
25_13	25	1782	1645.2	2567.17	**685.74**	2408.69
30_8	30	926	**2743.51**	> 3600	> 3600	> 3600
40_5	40	447	**1321.61**	> 3600	> 3600	> 3600

Instance			Resolution time (secs)				
ID	n	m	PC1	PC2	PC3	ISHIKAWA	LOGN-1
20_5	20	207	10.58	5.05	**4.27**	37.47	35.34
20_10	20	833	**90.28**	159.47	137.69	417.72	365.47
25_6	25	407	106.67	**80.17**	121.03	629.66	466.92
30_4	30	223	13.52	7.17	**7.03**	29.67	36.08
35_4	35	263	24.13	13.25	**11.2**	49.77	54.14
25_13	25	1782	**2311.09**	> 3600	> 3600	> 3600	> 3600
30_8	30	926	> 3600	> 3600	> 3600	> 3600	> 3600
40_5	40	447	> 3600	**914.27**	2053.97	> 3600	> 3600

Autocorr. seq.: Quadratization properties

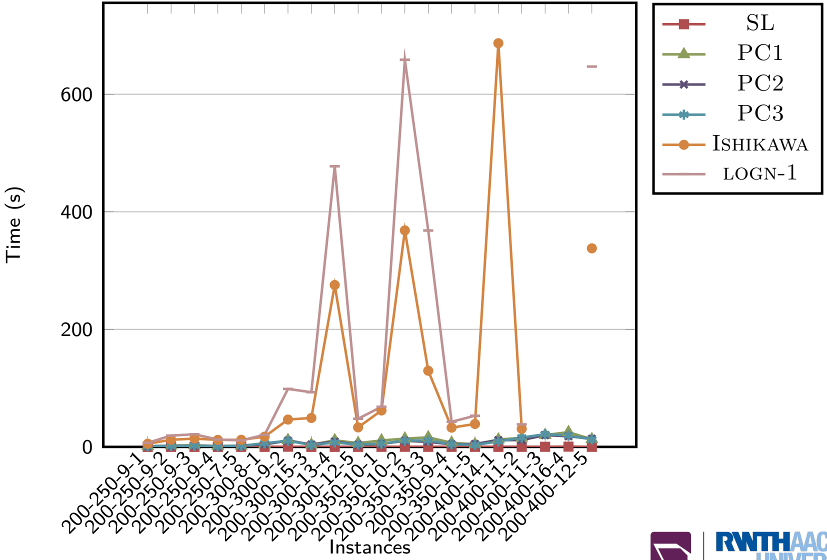
	Pairwise covers	Termwise
Number of y variables	less	more
Number of positive quadratic terms	less	more
Monomial interactions	✓	✗

What happens with random polynomials?

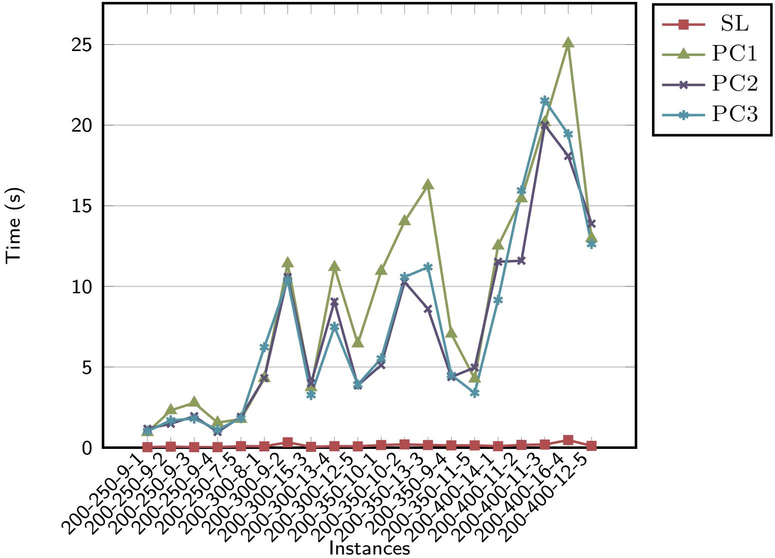
Generated as in (Buchheim & Rinaldi, 2007).

- Degree of each monomial chosen randomly (higher probability for lower degrees).
- The variables in monomials are randomly chosen, and the coefficients.
- Degrees vary between 7 and 17.

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
Monomial interactions	✓	✗

RWTH AACHEN UNIVERSITY

Summary of results

- Linearizations are in general solved faster than quadratizations, which is expected for CPLEX.

Summary of results

- Linearizations are in general solved faster than quadratizations, which is expected for CPLEX.
- For vision instances the behavior is different and rather surprising:

Summary of results

- Linearizations are in general solved faster than quadratizations, which is expected for CPLEX.
- For vision instances the behavior is different and rather surprising:
  - Pairwise covers are faster than SL with default CPLEX branch & cut.

Summary of results

- Linearizations are in general solved faster than quadratizations, which is expected for CPLEX.
- For vision instances the behavior is different and rather surprising:
  - Pairwise covers are faster than SL with default CPLEX branch & cut.
  - SL with 2-links is faster than pairwise covers.

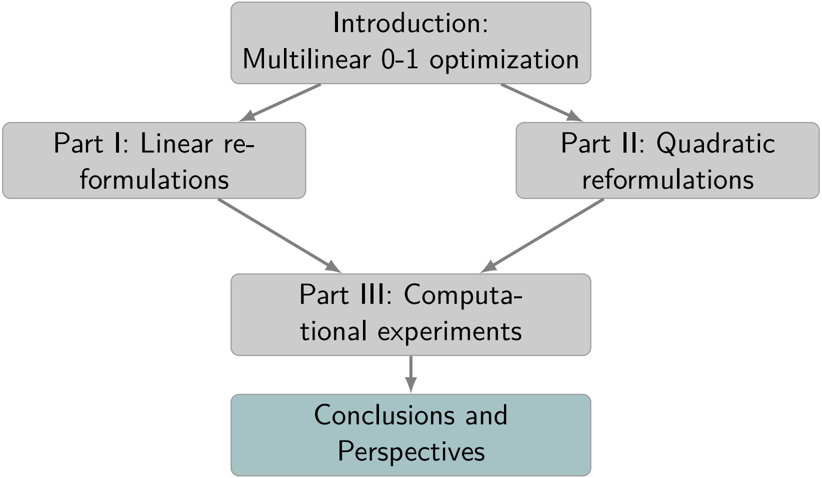
Summary of results

- Linearizations are in general solved faster than quadratizations, which is expected for CPLEX.
- For vision instances the behavior is different and rather surprising:
  - Pairwise covers are faster than SL with default CPLEX branch & cut.
  - SL with 2-links is faster than pairwise covers.
- For all instances, termwise quadratizations are slower than pairwise covers.

Summary of results

- Linearizations are in general solved faster than quadratizations, which is expected for CPLEX.
- For vision instances the behavior is different and rather surprising:
  - Pairwise covers are faster than SL with default CPLEX branch & cut.
  - SL with 2-links is faster than pairwise covers.
- For all instances, termwise quadratizations are slower than pairwise covers.
- Minimizing the number of auxiliary variables is not the only criterion to consider.

Overview


Introduction:
Multilinear 0-1 optimization
Part I: Linear re-
formulations
Part II: Quadratic
reformulations
Part III: Computa-
tional experiments
Conclusions and
Perspectives

Conclusions

Conclusions

Summary

Conclusions

Summary

- We considered linear and quadratic reformulations.

Conclusions

Summary

- We considered linear and quadratic reformulations.
- We derived several theoretical results (strong linear formulations, study of properties of quadratizations, small number of auxiliary variables...).

Conclusions

Summary

- We considered linear and quadratic reformulations.
- We derived several theoretical results (strong linear formulations, study of properties of quadratizations, small number of auxiliary variables...).
- We compared the reformulations computationally.

Conclusions

Summary

- We considered linear and quadratic reformulations.
- We derived several theoretical results (strong linear formulations, study of properties of quadratizations, small number of auxiliary variables...).
- We compared the reformulations computationally.
- Some properties to define good reformulations

Conclusions

Summary

- We considered linear and quadratic reformulations.
- We derived several theoretical results (strong linear formulations, study of properties of quadratizations, small number of auxiliary variables...).
- We compared the reformulations computationally.
- Some properties to define good reformulations
  - Exploiting structural properties (worked best for vision and physics problems).

Conclusions

Summary

- We considered linear and quadratic reformulations.
- We derived several theoretical results (strong linear formulations, study of properties of quadratizations, small number of auxiliary variables...).
- We compared the reformulations computationally.
- Some properties to define good reformulations
  - Exploiting structural properties (worked best for vision and physics problems).
  - For quadratizations, understand number of positive quadratic terms influence.

Open questions

Theoretical

Open questions

Theoretical

- Which one is stronger: SL or SL of a quadratization?

Open questions

Theoretical

- Which one is stronger: SL or SL of a quadratization?
- In general, understand better which properties define a "good" quadratization, theory and practice.

Open questions

Theoretical

- Which one is stronger: SL or SL of a quadratization?
- In general, understand better which properties define a “good” quadratization, theory and practice.
- ...

Open questions

Theoretical

- Which one is stronger: SL or SL of a quadratization?
- In general, understand better which properties define a "good" quadratization, theory and practice.
- ...

Experiments

Open questions

Theoretical

- Which one is stronger: SL or SL of a quadratization?
- In general, understand better which properties define a "good" quadratization, theory and practice.
- ...

Experiments

- Experiments are being re-tested using persistencies.

Open questions

Theoretical

- Which one is stronger: SL or SL of a quadratization?
- In general, understand better which properties define a "good" quadratization, theory and practice.
- ...

Experiments

- Experiments are being re-tested using persistencies.
- Repeat experiments with other solvers (convexification, SDP...).

RWTH AACHEN UNIVERSITY

Open questions

Theoretical

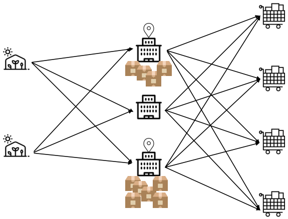
- Which one is stronger: SL or SL of a quadratization?
- In general, understand better which properties define a "good" quadratization, theory and practice.
- ...

Experiments

- Experiments are being re-tested using persistencies.
- Repeat experiments with other solvers (convexification, SDP...).
- ...

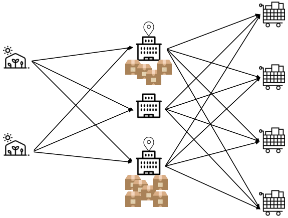
Open question: relaxing the multilinear assumption

Joint supply chain design & inventory management problem (Shen, Coullard, & Daskin, 2003; You & Grossmann, 2008)



Open question: relaxing the multilinear assumption

Joint supply chain design & inventory management problem (Shen, Coullard, & Daskin, 2003; You & Grossmann, 2008)



A pseudo-Boolean formulation (with constraints)

$$\min \sum_{j \in J} f_j x_j + \sum_{i \in I} \hat{d}_{ij} y_{ij} +$$

$$K_j \sqrt{\sum_{i \in I} \mu_i y_{ij}} + q \sqrt{\sum_{i \in I} \hat{\sigma}_i^2 y_{ij}}$$

$$\text{s.t.} \sum_{j \in J} y_{ij} = 1, \quad \forall i \in I$$

$$y_{ij} \leq x_j, \quad \forall i \in I, \forall j \in J$$

$$x_j \in \{0,1\}, y_{ij} \in \{0,1\} \quad \forall i \in I, \forall j \in J$$

Thank you for your attention!

rodriguez-heck@or.rwth-aachen.de

Bibliography I

Anthony, M., Boros, E., Crama, Y., & Gruber, A. (2016). Quadratization of symmetric pseudo-Boolean functions. *Discrete Applied Mathematics*, *203*, 1 – 12.

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

Balas, E. (1974). Disjunctive programming: properties of the convex hull of feasible points. *GSIA Management Science Research Report MSRR 348, Carnegie Mellon University (1974).*. (Published as invited paper in *Discrete Applied Mathematics* 89 (1998) 3–44)

Bernasconi, J. (1987). Low autocorrelation binary sequences: statistical mechanics and configuration space analysis. *J. Phys. France*, *48*(4), 559–567.

Bibliography II

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

Boros, E., & Hammer, P. L. (2002). Pseudo-Boolean optimization. *Discrete Applied Mathematics*, *123*(1), 155–225.

Boros, E., Hammer, P. L., Sun, X., & Tavares, G. (2008). A max-flow approach to improved lower bounds for quadratic unconstrained binary optimization (QUBO). *Discrete Optimization*, *5*(2), 501–529. (In Memory of George B. Dantzig)

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.

Bibliography III

Conforti, M., & Cornuéjols, G. (1995). Balanced 0,±1-matrices, bicoloring and total dual integrality. *Mathematical Programming*, *71*(3), 249–258.

Conforti, M., Cornuéjols, G., & Vušković, K. (2006). Balanced matrices. *Discrete Mathematics*, *306*(19–20), 2411–2437.

Crama, Y. (1993). Concave extensions for nonlinear 0–1 maximization problems. *Mathematical Programming*, *61*(1), 53–60.

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

Del Pia, A., & Khajavirad, A. (2016). *On balanced matrices and polynomial solvability of multilinear programs.* (Personal communication)

Bibliography IV

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 V

Hammer, P. L., Hansen, P., & Simeone, B. (1984). Roof duality, complementation and persistency in quadratic 0—1 optimization. *Mathematical Programming*, *28*(2), 121–155.

Hammer, P. L., Rosenberg, I., & Rudeanu, S. (1963). On the determination of the minima of pseudo-Boolean functions. *Studii si Cercetari Matematice*, *14*, 359–364. (in Romanian)

Hansen, P., & Simeone, B. (1986). Unimodular functions. *Discrete Applied Mathematics*, *14*(3), 269 – 281.

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.

Bibliography VI

Liers, F., Marinari, E., Pagacz, U., Ricci-Tersenghi, F., & Schmitz, V. (2010). A non-disordered glassy model with a tunable interaction range. *Journal of Statistical Mechanics: Theory and Experiment*(5), L05003.

Padberg, M. (1989). The boolean quadric polytope: some characteristics, facets and relatives. *Mathematical Programming*, *45*(1–3), 139–172.

Shen, Z.-J. M., Coullard, C., & Daskin, M. S. (2003). A joint location-inventory model. *Transportation Science*, *37*(1), 40-55.

You, F., & Grossmann, I. E. (2008). Mixed-integer nonlinear programming models and algorithms for large-scale supply chain design with stochastic inventory management. *Industrial & Engineering Chemistry Research*, *47*(20), 7802–7817.