A computational comparison of quadratizations for polynomial binary optimization problems

Elisabeth Rodríguez-Heck
joint work with
Endre Boros (Rutgers University), and
Yves Crama (University of Liège)

June 25, 2019

EURO 30th Conference, Dublin

The problem

Multilinear binary optimization

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

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

Example:

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

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

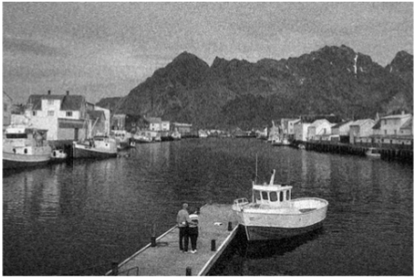
$$f(x) = \min\{g(x,y) : y \in \{0,1\}^m\} \quad \forall x \in \{0,1\}^n.$$

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).
- ...

Motivations to use quadratizations

Input: blurred image



Output: restored image

Image from the Corel database.

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)

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

such that $\tilde{x}_j = 1$ for $j \in O$ and $\tilde{x}_j = 0$ for $j \in Z$.

- Then, there exists an optimal solution x^* of (QP) such that $x^*_j = 1$ for $j \in O$ and $x^*_j = 0$ for $j \in Z$.

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 general, the set of variables fixed with the linearization is smaller as with the network flow algorithm.
- In computer vision, image restoration and related problems of up to *millions* of variables are efficiently solved using persistencies (usually not to optimality).
- How much persistencies can help depends on the size of the set of variables that one can fix.

What about linearizing directly?

[Advertising break]

EURO Doctoral Dissertation Award Session

Next session: TC-02: EDDA
Tuesday, 12:30-14:00
Moore Auditorium

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.

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 $y = 0$.

RWTH AACHEN UNIVERSITY

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.

RWTH AACHEN UNIVERSITY

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 3

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

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*.

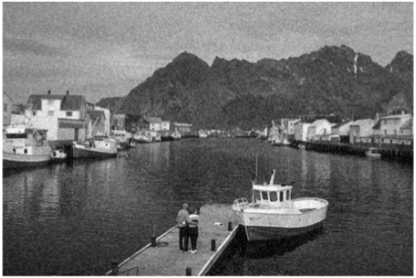
Computational experiments: Methods and Scope

Methods					
Linearizations	Quadratizations				
(Fortet, 1959)	*Pairw. cov.*			*Termwise*	
SL	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.

Instances: Vision

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

→

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

1 1	1 1	term: $10x_1x_2x_3x_4$
0 1	1 0	term: $40(1-x_1)x_2x_3(1-x_4)$
0 1	0 0	term: $20(1-x_1)x_2(1-x_3)(1-x_4)$

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

→

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

Base images:

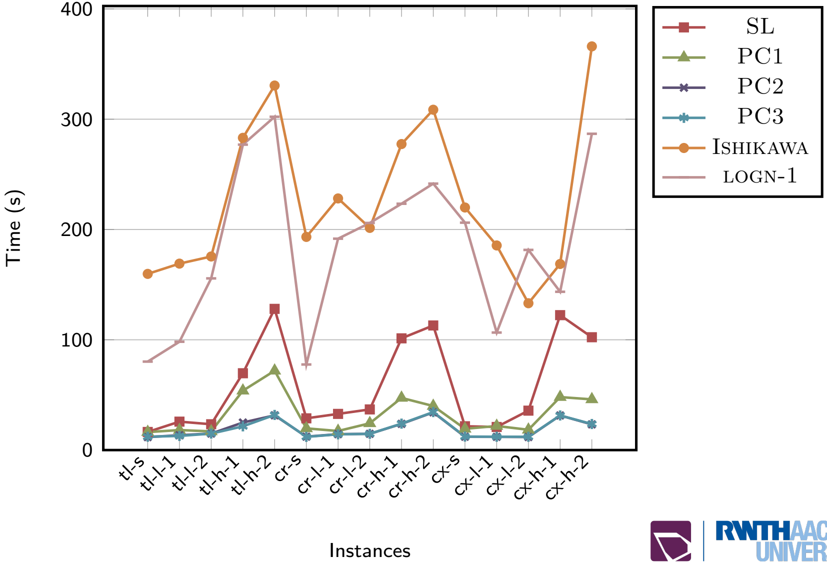
- top left rect. (tl)
- centre rect. (cr)
- cross (cx)

Perturbations:

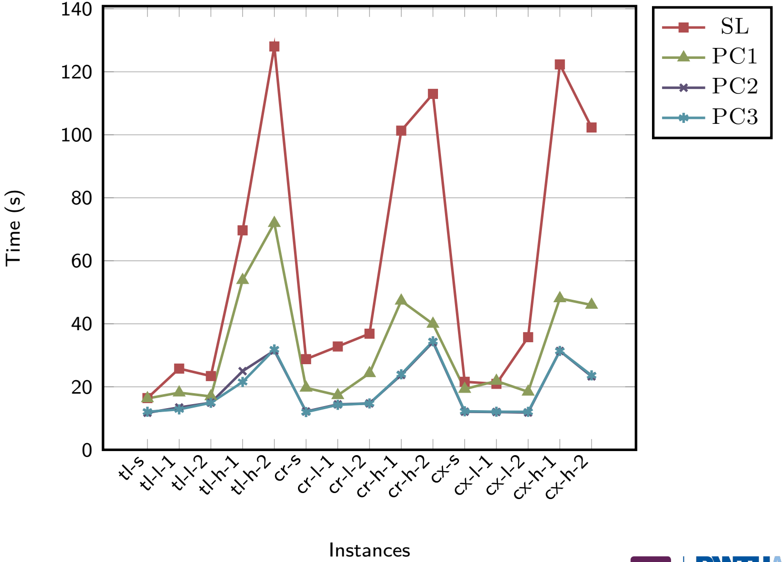
- none (n)
- low (l)
- high (h)

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

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



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



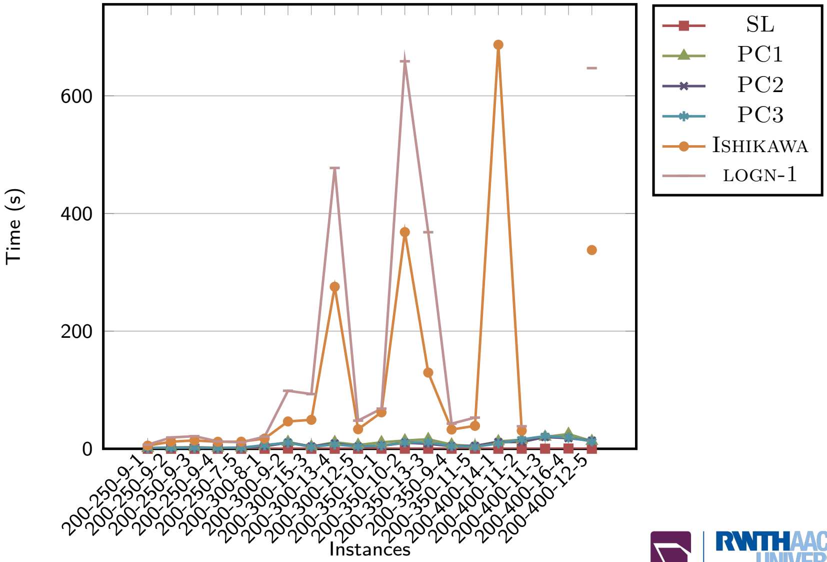
Vision: Quadratization properties

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

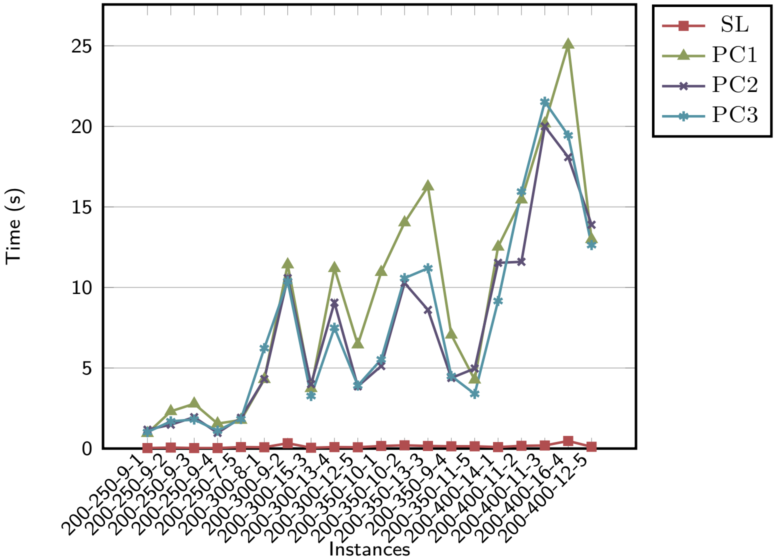
Instances

- 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 polynomials: all methods



Random polynomials: best methods



Random high degree: Quadratization properties

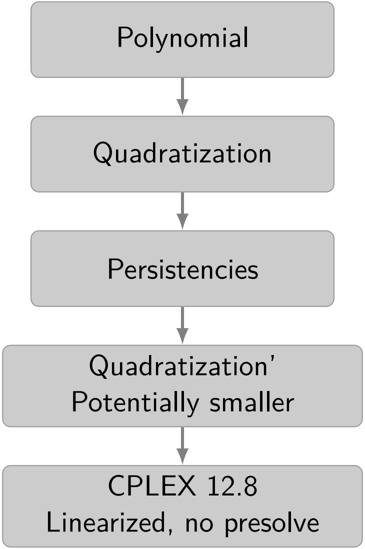
	Pairwise covers	Termwise
Number of y variables	**more**	**less**
Number of positive quadratic terms	less	more

- Small number of y variables might not be the best or only criterion to consider!

Current work

- Persistencies showed promising results in the computer vision community when solving image restoration problems through quadratization (usually not to optimality).
- Can persistencies help for solving quadratizations *exactly*?
- Open-source implementation of persistencies computation by Vladimir Kolmogorov `https://pub.ist.ac.at/~vnk/software.html`.

Current work



Persistencies: first results

Number of variables fixed:

Instance set	PC: fixed	Termwise: fixed
Vision	0 (variables)	0 (variables)
Autocorrelated sequences	1 (variable)	1 (variable)
Random (high degree)	0.2-6.11 (%)	6.08 - 16.91 (%)

Computing times for Random (high degree):

- Pairwise Covers times very similar.
- Results for Termwise.

Persistencies: first results

- Termwise: **Ishikawa** for positive monomials
- # variables: quadratization + linearization

Instance	# vbles	(% fix)	time	time (P)	nodes	nodes (P)
200-250-1-9	344	11.05	7.26	9.52	19	9
200-250-2-9	342	14.92	25.75	23.79	1 032	1 240
200-250-3-9	339	11.2	34.75	26.88	1 807	396
200-250-4-9	340	16.77	18.08	16.27	596	133
200-250-5-7	340	15.88	20.6	15.02	250	319
200-300-1-8	369	9.21	97.68	117.21	9 835	14 155
200-300-2-9	390	7.18	190.43	396.93	11 290	42 100
200-300-3-15	372	8.06	381.43	259.97	19 477	11 121
200-300-4-13	378	6.08	> 3600	1185.24	246 400	59 849
200-300-5-12	373	7.78	59.67	55.65	1 384	1 485

Persistencies: first results

- Termwise: $\lceil \log(n) \rceil - 1$ for positive monomials
- # variables: quadratization + linearization

Instance	# vbles	(% fixed)	time	time (P)	nodes	nodes (P)
200-250-1-9	341	11.14	12.47	11.64	28	17
200-250-2-9	339	15.04	53.7	70.93	6 616	7 784
200-250-3-9	334	11.37	46.32	59.61	2 425	3 313
200-250-4-9	337	16.91	25.2	22.27	1 025	764
200-250-5-7	338	15.98	21.48	33.12	333	2 800
200-300-1-8	367	9.26	170.54	209.5	17 475	26 977
200-300-2-9	385	7.01	309.82	2556.18	27 977	332 356
200-300-3-15	362	8.28	794.01	634.69	54 318	49 784
200-300-4-13	369	5.42	> 3600	> 3600	219 930	196 152
200-300-5-12	369	7.86	615.17	544	45 031	47 499

Conclusions

Summary

- New compact quadratizations for the positive monomial.
- Proof of the lower bound on the number of auxiliary variables.
- First experiments
  - Small number of auxiliary variables might not be the best or only criterion to define good quadratizations.
  - Persistencies results difficult to evaluate.

Perspectives

- Persistencies experiments:
  - Test other solvers, other parameter settings
  - Understand is forcing persistencies makes us look for a different optimal solution that is more difficult to find.
- Better understand properties defining good quadratizations.

Bibliography I

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)

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

Bibliography II

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).

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

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.

Bibliography III

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.