An Exact Penalty Method over Discrete Sets

Nicolò Gusmeroli

joint work with Angelika Wiegele

TU Dortmund

nicolo.gusmeroli@math.tu-dortmund.de

Novmeber 24, 2020

ESR Days - online event

technische universität
dortmund

BiqBin

Project:

"High-Performance solver for Binary Quadratic Problems"

BiqBin

Project:

"High-Performance solver for Binary Quadratic Problems"

Given a quadratic objective function $f(x) = c^\top x + x^\top F x$, where $c \in \mathbb{R}^n$ and $F \in \mathbb{R}^{n \times n}$, and the equality constraints $Ax = b$, where $A \in \mathbb{Z}^{m \times n}$ and $b \in \mathbb{Z}^m$, we want to solve

$$f^* := \min \left\{ f(x) \colon Ax = b,\ x \in \{-1, 1\}^n \right\}.$$

BiqBin

Project:

"High-Performance solver for Binary Quadratic Problems"

Given a quadratic objective function $f(x) = c^\top x + x^\top Fx$, where $c \in \mathbb{R}^n$ and $F \in \mathbb{R}^{n\times n}$, and the equality constraints $Ax = b$, where $A \in \mathbb{Z}^{m\times n}$ and $b \in \mathbb{Z}^m$, we want to solve

$$f^* := \min \left\{ f(x) \colon Ax = b,\ x \in \{-1, 1\}^n \right\}.$$

⇒ novel solver **BiqBin**

1. exact penalty method over discrete sets `EXPEDIS` which reformulates the input problem into an equivalent max-cut instance
2. improved semidefinite-based branch-and-bound algorithm for max-cut

⇒ use of a high-performance computer located in Slovenia

⇒ web application and more info at `http://www.biqbin.eu/`

Exact Penalty Method Over Discrete Sets

Consider the original problem

$$f^* := \min \Big\{ f(x) \colon Ax = b,\ x \in \{-1, 1\}^n \Big\},$$

we derive an equivalent max-cut instance.

Exact Penalty Method Over Discrete Sets

Consider the original problem

$$f^* := \min \left\{ f(x) \colon Ax = b,\ x \in \{-1, 1\}^n \right\},$$

we derive an equivalent max-cut instance.

We use **EXPEDIS**, an EXact PEnalty method over DIscrete Sets:

- initial problem
- computation of bounds by using some SDP-relaxations
- definition of the *penalty* and the *threshold* parameters
- Lagrangian approach penalizing the equality constraints
- reformulation as *max-cut*

Penalization

Consider the original problem

$$f^* \coloneqq \min \left\{ f(x) \colon Ax = b,\ x \in \{-1, 1\}^n \right\},$$

we want to penalize the linear constraints.

Penalization

Consider the original problem

$$f^* := \min\left\{f(x)\colon Ax = b,\ x \in \{-1,1\}^n\right\},$$

we want to penalize the linear constraints.

We use a Lagrangian approach on the equality constraint $Ax = b$, thus, given a suitable penalty parameter σ, we obtain the penalized function

$$h(x) = f(x) + \sigma\|Ax - b\|^2$$

and the penalized problem

$$h^* = \min\left\{h(x)\colon\ x \in \{-1,1\}^n\right\}.$$

Main theorem

For ease the notation we partition the set $\{-1, 1\}^n$ into

$$\Delta = \{x \in \{-1, 1\}^n \colon Ax = b\} \quad \text{and } \Delta^c = \{-1, 1\}^n \backslash \Delta.$$

Theorem

Let f^* and h^* be the optimal values of the original and the penalized problem, respectively. Furthermore, assume we have a threshold parameter ρ and a penalty parameter σ, satisfying the following conditions:

- *no feasible solution of the original problem has value greater than ρ,*
- *for every $x \in \Delta^c$, we have $h(x) > \rho$.*

Main theorem

For ease the notation we partition the set $\{-1,1\}^n$ into

$$\Delta = \{x \in \{-1,1\}^n \colon Ax = b\} \quad \text{and } \Delta^c = \{-1,1\}^n \setminus \Delta.$$

Theorem

Let f^* and h^* be the optimal values of the original and the penalized problem, respectively. Furthermore, assume we have a threshold parameter ρ and a penalty parameter σ, satisfying the following conditions:

- *no feasible solution of the original problem has value greater than ρ,*
- *for every $x \in \Delta^c$, we have $h(x) > \rho$.*

If $f^* < +\infty$, then $h^* = f^*$. Moreover the original problem is infeasible if and only if $h^* > \rho$.

Proof

By integrality of A, b and x we have that $\|Ax - b\|^2 \in \mathbb{Z}$.

Recalling the definitions:

$$f(x) = c^\top x + x^\top Fx \qquad\qquad h(x) = c^\top x + x^\top Fx + \sigma\|Ax - b\|^2$$

It follows $\|Ax - b\|^2 \begin{cases} = 0 & \implies h(x) = f(x) + 0 = f(x) \leqslant \rho \quad \text{if } x \in \Delta \\ \geqslant 1 & \implies h(x) \geqslant f(x) + 1 \cdot \sigma > \rho \quad \text{if } x \in \Delta^c \end{cases}$

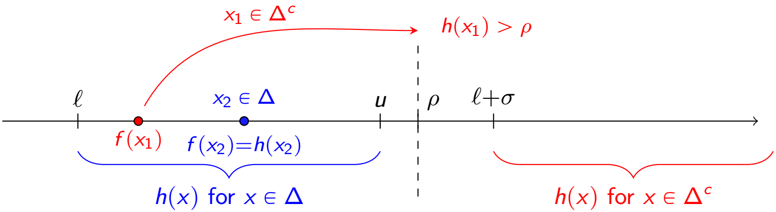
Proof

By integrality of A, b and x we have that $\|Ax - b\|^2 \in \mathbb{Z}$.

Recalling the definitions:

$$f(x) = c^\top x + x^\top Fx \qquad h(x) = c^\top x + x^\top Fx + \sigma\|Ax - b\|^2$$

It follows $\|Ax - b\|^2 \begin{cases} = 0 & \implies h(x) = f(x) + 0 = f(x) \leqslant \rho & \text{if } x \in \Delta \\ \geqslant 1 & \implies h(x) \geqslant f(x) + 1 \cdot \sigma > \rho & \text{if } x \in \Delta^c \end{cases}$



Final Reformulation

Max-cut instance

- expand the norm $\|Ax - b\|^2$ as $(Ax - b)^\top (Ax - b)$
- increase the dimension by one and fix first variable to 1
- computation of $Q \in \mathbb{R}^{(n+1)\times(n+1)}$ depending on A, b, c, F and ρ

Final Reformulation

Max-cut instance

- expand the norm $\|Ax - b\|^2$ as $(Ax - b)^\top (Ax - b)$
- increase the dimension by one and fix first variable to 1
- computation of $Q \in \mathbb{R}^{(n+1)\times(n+1)}$ depending on A, b, c, F and ρ

$$\implies h^* = \min \left\{ \bar{x}^\top Q \bar{x} \colon \bar{x} \in \{-1, 1\}^{n+1},\ \bar{x}_0 = 1 \right\}$$

The adjacency matrix of the resulting graph is

$$\mathcal{A}_{ij} = \begin{cases} 0 & \text{if } i = j \\ 2c_j - 4\sigma \left(A_{\bullet,j}\right)^\top b & \text{if } i = 0 \text{ and } i \neq j \\ 2c_i - 4\sigma \left(A_{\bullet,i}\right)^\top b & \text{if } j = 0 \text{ and } j \neq i \\ 4F_{i,j} + 4\sigma \left(A_{\bullet,j}\right)^\top A_{\bullet,i} & \text{if } 1 \leqslant i,j \leqslant n \text{ and } i \neq j \end{cases}$$

Considerations on Parameters

Consideration

We transform a binary quadratic problem with equality constraints into a max-cut instance. The weights of the edges depend on the parameters $A, b, c, , F$ and by the penalizer σ.

Considerations on Parameters

Consideration

We transform a binary quadratic problem with equality constraints into a max-cut instance. The weights of the edges depend on the parameters $A, b, c, , F$ and by the penalizer σ.

The choice of the penalty parameter has significant influence on the tractability of the problem.

We want to keep its value as small as possible!

Smallest penalty parameter

Upper and lower bounds

Considering the lower and the upper bounds

$$\ell^* = \min\{f(x) \colon x \in \Delta^c\} \quad \text{and} \quad u^* = \max\{f(x) \colon x \in \Delta\},$$

we can define the following parameters:

Smallest penalty parameter

Upper and lower bounds

Considering the lower and the upper bounds

$$\ell^* = \min\{f(x) \colon x \in \Delta^c\} \quad \text{and} \quad u^* = \max\{f(x) \colon x \in \Delta\},$$

we can define the following parameters:

$$\rho^* = u^* \quad \text{and} \quad \sigma^* = u^* - \ell^* + \epsilon.$$

Smallest penalty parameter

Upper and lower bounds

Considering the lower and the upper bounds

$$\ell^* = \min\{f(x) \colon x \in \Delta^c\} \quad \text{and} \quad u^* = \max\{f(x) \colon x \in \Delta\},$$

we can define the following parameters:

$$\rho^* = u^* \quad \text{and} \quad \sigma^* = u^* - \ell^* + \epsilon.$$

Theorem

The penalty parameter $\sigma^* = u^* - \ell^* + \epsilon$ is the smallest possible, such that the assumptions of the main theorem hold.

Efficiently computable penalty parameter

Problem

Solving

$$\ell^* = \min\{f(x)\colon x \in \Delta^c\} \quad \text{and} \quad u^* = \max\{f(x)\colon x \in \Delta\},$$

is in general as difficult as solving our original problem.

Efficiently computable penalty parameter

Problem

Solving

$$\ell^* = \min\{f(x) \colon x \in \Delta^c\} \quad \text{and} \quad u^* = \max\{f(x) \colon x \in \Delta\},$$

is in general as difficult as solving our original problem.

Thus we want to derive some penalty parameter which

- is relatively small and
- can be computed quickly.

Theorem

Given any pair of bounds (ℓ, u) such that $\ell \leqslant \ell^*$ and $u \geqslant u^*$. Then it follows that the threshold $\rho = u$ and the penalty parameter $\sigma = u - \ell + \epsilon$ satisfy the assumptions of the main theorem.

Previous work

Bounds and parameters from Lasserre

As far as we know the only work in this direction was of Lasserre: given

$$\hat{\ell}(\hat{u}) = \min(\max)\left\{\langle F, X\rangle + c^\top x \colon \begin{pmatrix} 1 & x^\top \\ x & X \end{pmatrix} \succcurlyeq 0, X_{ii} = 1\right\},$$

he derived as parameters $\hat{\rho} = \max\{|\hat{\ell}|, |\hat{u}|\}$ and $\hat{\sigma} = 2\hat{\rho} + 1$.

Previous work

Bounds and parameters from Lasserre

As far as we know the only work in this direction was of Lasserre: given

$$\hat{\ell}(\hat{u}) = \min(\max)\left\{\langle F, X\rangle + c^\top x \colon \begin{Bmatrix} 1 & x^\top \\ x & X \end{Bmatrix} \succcurlyeq 0, X_{ii} = 1\right\},$$

he derived as parameters $\hat{\rho} = \max\{|\hat{\ell}|, |\hat{u}|\}$ and $\hat{\sigma} = 2\hat{\rho} + 1$.

Our bounds and parameters

Let $M = [b, -A]$ and $Y = \begin{bmatrix} 1 & x^\top \\ x & X \end{bmatrix}$, we calculate the bounds

$$u' = \max\left\{c^\top x + \mathrm{tr}(X^\top F) \mid Y \succcurlyeq 0,\ diag(X) = e,\ MY = \mathbf{0}\right\}$$
$$l' = \min\left\{c^\top x + \mathrm{tr}(X^\top F) \mid Y \succcurlyeq 0,\ diag(X) = e,\ X \in \mathrm{MET}\right\},$$

where MET is the set of triangle and some 5-clique inequalities, i.e., cutting planes which strengthen the relaxation.

Thus $\rho' = u'$ and $\sigma' = u' - \ell' + \epsilon$.

Comparisons

Theorem

From the previous definitions, it follows easily

$$\hat{\rho} \geqslant \rho' \quad \textit{and} \quad \hat{\sigma} > \sigma'.$$

Comparisons

Theorem

From the previous definitions, it follows easily

$$\hat{\rho} \geqslant \rho' \quad \textit{and} \quad \hat{\sigma} > \sigma'.$$

Considerations

We considered thousands of instances (both randomly generated and from the max k-cluster problem).

- the penalty parameter σ' is in general 20% of $\hat{\sigma}$
- computing σ' takes less than 2 minutes, while $\hat{\sigma}$ less than 30 seconds
- by using σ' almost 70% of the instances were solved within 1.5 hours, while by using $\hat{\sigma}$ less than 45%

Comparisons

Theorem

From the previous definitions, it follows easily

$$\hat{\rho} \geqslant \rho' \quad \textit{and} \quad \hat{\sigma} > \sigma'.$$

Considerations

We considered thousands of instances (both randomly generated and from the max k-cluster problem).

- the penalty parameter σ' is in general 20% of $\hat{\sigma}$
- computing σ' takes less than 2 minutes, while $\hat{\sigma}$ less than 30 seconds
- by using σ' almost 70% of the instances were solved within 1.5 hours, while by using $\hat{\sigma}$ less than 45%

These first computational results validated our intuitions, thus we choose σ' as penalty parameter. We compare the solution of the max-cut instance (by using BiqMac) with the solution of the original problem (by using commercial solvers).

Results on Randomly Generated Instances

Randomly Generated Instances

We consider randomly generated instances: the parameters A, b, c and F are random, from different sets. We consider instances with 80 and 100 variables and up to 15 constraints.

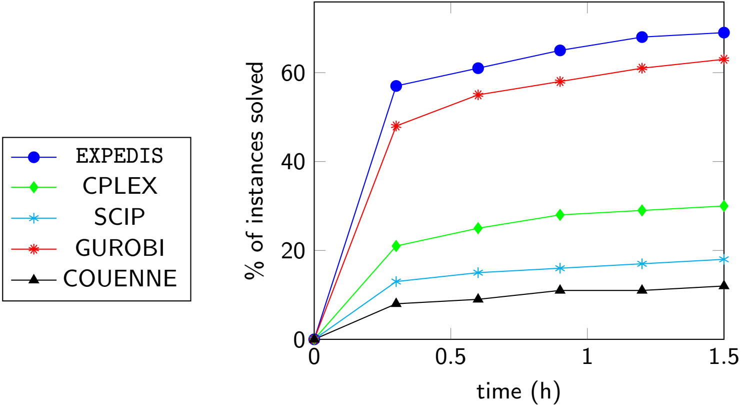


Figure: Performance profile of the different solvers on all the randomly generated

Future work

- improve the bounds by using different relaxations
- test on different problems
- use some specific approach for detecting possible infeasibility
- find a way to penalize also some quadratic constraints