A Conic Relaxation Approach for Semi-Integer Problems arising from Tree Breeding

Sepuluh Nopember Institute of Technology * SAFARINA Sena
01704970 Tokyo Institute of Technology YAMASHITA Makoto

1. Introduction

In many optimization problems, a decision variable x is mostly confined in continuous decision variables $l \le x \le u$ for lower l and upper u bounds, or in discrete decision variables $x \in \mathbb{Z}[a, b] := \{a, a+1, a+2, \dots, b-1, b\}$. However, in many real-world application, there are some cases that the decision variable should be confined in a disconnected set $x \in \{l\} \cup \mathbb{Z}[a, b]$, that is, the feasible solution is either $\{l\}$ or an integer between a and b inclusive. However, such semi-integer variables are usually more difficult to solve than continuous variables.

In this research, we apply a cone decomposition method proposed in [3] to reduce a computation time for solving a semi-integer problem arising from tree breeding below:

$$\begin{array}{rll} \max_{x \in \mathbb{R}^Z} & : & \boldsymbol{g}^T \boldsymbol{x} \\ \text{s.t.} & : & \boldsymbol{e}^T \boldsymbol{x} = N \\ & & \boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} \le 2\theta N^2 \\ & & x_i \in \{l_i\} \cup \mathbb{Z}[a_i, b_i] \\ & & (i = 1, \dots, Z). \end{array} \tag{1}$$

where N is a given integer and θ is a threshold. The vector $\boldsymbol{g} \in \mathbb{R}^Z$ is a given vector, and $\boldsymbol{e}$ is a vector of all ones. It is known that the numerator relationship matrix $\boldsymbol{A}$ is always positive definite. For more details on $\boldsymbol{A}$, refer to [4].

2. CDM

Due the positive definiteness of $\boldsymbol{A}$, there is a matrix $\hat{\boldsymbol{B}}$ such that $\boldsymbol{A} = \hat{\boldsymbol{B}} \hat{\boldsymbol{B}}^T$, therefore, we can convert (1) into a mixed-integer SOCP (MI-SOCP) as follows.

$$\begin{array}{rll} \max & : & \boldsymbol{g}^T \boldsymbol{x} \\ \text{s.t.} & : & \boldsymbol{e}^T \boldsymbol{x} = N \\ & & \|\hat{\boldsymbol{B}} \boldsymbol{x}\| \le \sqrt{2\theta} N \\ & & (a_i - l_i) y_i \le x_i - l_i \le (b_i - l_i) y_i \\ & & x_i \in \mathbb{Z}, y_i \in \{0, 1\} \ (i = 1, \dots, Z) \end{array} \tag{2}$$

Cone decomposition method (CDM) was proposed by [3] to solve MI-SOCP problem combining a cutting plane method and a mathematical strucuture of second-order cones. They showed that CDM with sparse matrix can solve MI-SOCP problem effectively.

According to (2), we can consider another equivalent form of the quadratic constraint $\|\hat{\boldsymbol{B}} \boldsymbol{x}\| \le \sqrt{2\theta} N$ in (2). Introducing $\boldsymbol{v} = \boldsymbol{A}^{-1} \boldsymbol{x}$ and using the matrix $\boldsymbol{B}$ such that $\boldsymbol{A}^{-1} = \boldsymbol{B}^T \boldsymbol{B}$ and letting $\boldsymbol{b}_i^T$ be the ith row of $\boldsymbol{B}$, the cone decomposition of the quadratic constraint can be written as follows.

$$\begin{array}{rlcl} & \boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} \le 2\theta N^2 & \Leftrightarrow & \boldsymbol{v}^T \boldsymbol{A}^{-1} \boldsymbol{v} \le 2\theta N^2 \\ \Leftrightarrow & \|\boldsymbol{B} \boldsymbol{v}\|^2 \le 2\theta N^2 & \Leftrightarrow & \sum_{i=1}^Z (\boldsymbol{b}_i^T \boldsymbol{v})^2 \le 2\theta N^2 \\ \Leftrightarrow & \multicolumn{3}{l}{(\boldsymbol{b}_i^T \boldsymbol{v})^2 \le w_i (i = 1, \dots, Z), \sum_{i=1}^Z w_i \le 2\theta N^2} \end{array}$$

Hence, we can reformulate (2) into

$$\begin{array}{rll} \max & : & \boldsymbol{g}^T \boldsymbol{x} \\ \text{s.t.} & : & \boldsymbol{e}^T \boldsymbol{x} = N \\ & & \boldsymbol{x} = \boldsymbol{A}^{-1} \boldsymbol{v} \\ & & \sum_{i=1}^Z w_i \le 2\theta N^2 \\ & & (\boldsymbol{b}_i^T \boldsymbol{v})^2 \le w_i \\ & & (a_i - l_i) y_i \le x_i - l_i \le (b_i - l_i) y_i \\ & & x_i \in \mathbb{Z}, y_i \in \{0, 1\}, w_i \ge 0 \\ & & \boldsymbol{v} \in \mathbb{R}^Z \end{array} \tag{3}$$

The quadratic constraints $(\boldsymbol{b}_i^T \boldsymbol{v})^2 \le w_i$ cannot be directly handled by MILP solvers. Thus, in CDM, We first ignore the quadratic constraints, and we obtain some solution $\bar{\boldsymbol{v}}$ by an MILP solver, but $\bar{\boldsymbol{v}}$ often violates the quadratic constraints. To remove such solution $\bar{\boldsymbol{v}}$, we we generate a geometric cut (GC) using the structure of $(\boldsymbol{b}_i^T \boldsymbol{v})^2 \le w_i$ and add the cut to the MILP. CDM adds GCs iteratively, and it was shown in [3] that CDM can find the optimal solution of (1) in a finite number of iterations.

In this research, we also take into consideration an outer-approximation cut (OAC) $(\boldsymbol{A} \bar{\boldsymbol{x}})^T \boldsymbol{x} \le \sqrt{2\theta} N \sqrt{\bar{\boldsymbol{x}}^T \boldsymbol{A} \bar{\boldsymbol{x}}}$ proposed in [1] for the original quadratic constraint $\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} \le 2\theta N^2$. We can add OAC and GC at the same time in each iteration.

Table 1: Comparison among the solvers

Model	t_{tot} (s)	t_s (s)	$\boldsymbol{g}^T\boldsymbol{x}$	$\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}$
		$N_p = 20$		
CPLEX[OA]	1.64	1.26	612070.00	150000.00
CPLEX[no]	0.04	0.03	612070.00	150000.00
CBC[OA]	13.90	12.59	612070.00	150000.00
CBC[no]	0.04	0.03	612070.00	150000.00
GLPK[OA]	8.77	7.64	612070.00	150000.00
GLPK[no]	0.03	0.01	612070.00	150000.00
		$N_p = 829$		
CPLEX[OA]	13.58	7.64	1217101.61	1289211.00
CPLEX[no]	11.23	7.87	1216337.20	1280302.99
CBC[OA]	81.68	76.74	1218160.18	1291684.00
CBC[no]	420.87	413.40	1218206.04	1291326.00
GLPK[OA]	711.14	705.96	1218095.75	1291569.50
GLPK[no]	81.54	76.67	1217821.13	1291460.00

3. Numerical Result

Numerical experiments were conducted to generate solution through cone decomposition method (CDM) and its improvement by outer approximation (CDM-OA). We implemented the method by using `Julia 1.5.1` and setting `CPLEX v0.6.6, CBC v0.7.0`, and `GLPK v0.13.0` as an MILP solver. We then compared their performance with `CPLEX` of the MI-SOCP problem (2). All methods were executed on a 64-bit Windows 10 PC with Core i7-7660 CPU (2.50 GHz) and 8 GB memory space. The data were generated from [2]. The full size of the test instances are $Z = 43115$. We set parameter $N = 3000$ and $2\theta N^2 = 1291680$, and as a stopping criterion for `CPLEX`, we used gap 5%. The computation time was limited to 3 hours for each execution.

Table 1 shows the comparison of different selected candidates and solvers. In the first stage, we reduced the number of parents to select the candidates based on the number of selected parents N_p. We generated different sizes, but here we only shows the smallest $N_p = 20$ and the largest number of selected parents $N_p = 829$. The first column shows the name of MILP solvers, and [OA] indicates outer approximation implementation while [no] means without OA. The second and third columns shows the total time (t_{tot}) and the required time by the solvers (t_s), respectively. Finally, the generated objective values $\boldsymbol{g}^T\boldsymbol{x}$ are presented in the fourth column; and remind that $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}$ in the last column should satisfy $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} \leq 2\theta N^2$.

From Table 1, we observe that all solutions of the smallest selected candidates satisfy $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} \leq 2\theta N^2$ and `GLPK` with no outer approximation method requires the shortest computation time comparing to other methods. However, when the number of selected candidates is increasing, `CPLEX` without outer approximation turns to be fastest. Different with others, the solution by `CBC[OA]` failed to satisfy the quadratic constraint $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} \leq 2\theta N^2$ slightly.

4. Conclusion and Future Work

We compare the performance of CDM with and without outer approximation employing different MILP solvers. As a future work, we will implement other cutting method to improve CDM performance.

Reference

[1] T. Mullin and P. Belotti. Using branch-and-bound algorithms to optimize selection of a fixed size breeding population under a relatedness constraint. *Tree genetics & genomes*, 12(1):4, 2016.

[2] T. Mullin and T. Persson. Assembling optimum breeding populations for the Swedish Scots pine breeding program. *Arbetsrapport från Skogforsk*, pages 951-2017, 2017.

[3] S. Safarina, T. J. Mullin, and M. Yamashita. Polyhedral-based methods for mixed-integer socp in tree breeding. *Journal of the Operations Research Society of Japan*, 62(4):133–151, 2019.

[4] M. Yamashita, T. J. Mullin, and S. Safarina. An efficient second-order cone programming approach for optimal selection in tree breeding. *Optimization Letters*, 12:1683-1697, 2018.