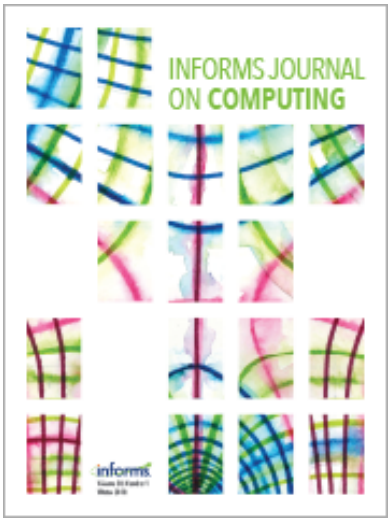


INFORMS Journal on Computing

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

A Dual Bounding Framework Through Cost Splitting for Binary Quadratic Optimization

Mahdis Bayani, Borzou Rostami, Yossiri Adulyasak, Louis-Martin Rousseau

To cite this article:
Mahdis Bayani, Borzou Rostami, Yossiri Adulyasak, Louis-Martin Rousseau (2024) A Dual Bounding Framework Through Cost Splitting for Binary Quadratic Optimization. INFORMS Journal on Computing

Published online in Articles in Advance 12 Mar 2024

. https://doi.org/10.1287/ijoc.2021.0186

https://pubsonline.informs.org/journal/ijoc

INFORMS JOURNAL ON COMPUTING
Articles in Advance, pp. 1–21
ISSN 1091-9856 (print), ISSN 1526-5528 (online)

A Dual Bounding Framework Through Cost Splitting for Binary Quadratic Optimization

Mahdis Bayani,[a,b,c,*] Borzou Rostami,[d] Yossiri Adulyasak,[c,e] Louis-Martin Rousseau[a,b]

[a] Polytechnique Montréal, Montreal, Quebec H3T 1J4, Canada; [b] CIRRELT, Université de Montréal, Montreal, Quebec H3T 1J4, Canada; [c] GERAD, Montreal, Quebec H3T 2A7, Canada; [d] Alberta School of Business, University of Alberta, Edmonton, Alberta T6G 2R3, Canada; [e] HEC Montréal, Montreal, Quebec H3T 2A7, Canada

*Corresponding author

Contact: mahdis.bayani@polymtl.ca, https://orcid.org/0000-0002-5047-3194 (MB); borzou@ualberta.ca, https://orcid.org/0000-0002-2610-1622 (BR); yossiri.adulyasak@hec.ca, https://orcid.org/0000-0002-6996-0742 (YA); louis-martin.rousseau@cirrelt.net, https://orcid.org/0000-0001-6949-6014 (L-MR)

Received: July 20, 2021
Revised: August 18, 2022; April 11, 2023; August 12, 2023; November 15, 2023; December 18, 2023; December 30, 2023
Accepted: January 4, 2024
Published Online in Articles in Advance: March 12, 2024

https://doi.org/10.1287/ijoc.2021.0186

Abstract. Binary quadratic programming (BQP) is a class of combinatorial optimization problems comprising binary variables, quadratic objective functions, and linear/nonlinear constraints. This paper examines a unified framework to reformulate a BQP problem with linear constraints to a new BQP with an exponential number of variables defined on a graph. This framework relies on the concept of stars in the graph to split the quadratic costs into adjacent and nonadjacent components indicating in-star and out-of-star interactions. We exploit the star-based structure of the new reformulation to develop a decomposition-based column generation algorithm. In our computational experiments, we evaluate the performance of our methodology on different applications with different quadratic structures. The quadratic component of the problem is dealt with in the column generation master problem and its subproblem. Results indicate the superiority of the framework over one of the state-of-the-art solvers, GUROBI, when applied to various benchmark reformulations with adjacent-only or sparse quadratic cost matrices. The framework outperforms GUROBI in terms of both dual bound and computational time in almost all instances.

History: Accepted by Andrea Lodi, Area Editor for Design & Analysis of Algorithms—Discrete.
Funding: The authors thank the Mitacs Accelerate Program for providing funding for this project. In addition, B. Rostami gratefully acknowledges the funding provided by the Canadian Natural Sciences and Engineering Research Council (NSERC) under a Discovery Grant [Grant RGPIN-2020-05395].
Supplemental Material: The online appendix is available at https://doi.org/10.1287/ijoc.2021.0186.

Keywords: binary quadratic programming • combinatorial optimization • column generation • semi-assignment problem • multiple object tracking problem

1. Introduction

Binary quadratic programming (BQP) is a large class of combinatorial optimization problems that arise from modeling real-life applications, for example, in management, engineering, logistics, and network design (Punnen et al. 2019). Given graph $G = (V, E)$ with node set $V = \{1, 2, \ldots, |V|\}$ and edge set $E = \{1, 2, \ldots, m\}$, a BQP problem with linear constraints on graph G can be specified using a quadratic cost matrix $\boldsymbol{q} \in \mathbb{R}^{m \times m}$ and a linear cost vector $\boldsymbol{c} \in \mathbb{R}^m$ and is formulated as follows:

$$\text{BQP}: \quad \min \quad \sum_{e \in E} c_e x_e + \sum_{(e,f) \in \mathcal{E}} q_{ef} x_e x_f \qquad \text{s.t.} \quad \boldsymbol{x} \in X, \tag{1}$$

where $X \subseteq \{0,1\}^m$ is the set of feasible binary vectors and $\mathcal{E} = \{(e,f) \in E \times E : e < f\}$. Without loss of generality, we only consider the symmetric quadratic costs in the objective function as one can replace q_{ef} and q_{fe} by their average $(q_{ef} + q_{fe})/2$ for the case of asymmetric quadratic costs. Note that the matrices and vectors are shown in bold text throughout the paper.

Many quadratic combinatorial optimization problems can be naturally formulated in this fashion. Some important examples include the quadratic assignment problem (Çela 2013), the quadratic knapsack problem (Pisinger 2007), the quadratic traveling salesman problem (Fischer 2014, Rostami et al. 2016, Punnen et al. 2017), the quadratic shortest path problem (Hu and Sotirov 2018, Rostami et al. 2018), the quadratic spanning tree problem (Assad and Xu 1992,

Rostami and Malucelli 2015, Pereira and da Cunha 2020), and the quadratic set covering problem (Escoffier and Hammer 2007, Punnen et al. 2019).

The main difficulty of solving BQP stems from the quadratic structure of the objective function rather than the combinatorial nature of the problem. For example, when $X = \{0,1\}^m$, problem (1) is equivalent to unconstrained quadratic binary optimization and hence to the max-cut problem, which is NP-hard (Barahona 1983). In general, BQP problems are NP-hard, even if the linear optimization problem over the same feasible set is tractable. This is the case for many BQP problems, including the quadratic spanning tree problem (Assad and Xu 1992) and the quadratic assignment problem (Çela 2013), although their linear counterparts are polynomially solvable.

One way to deal with the challenges of BQP is to identify patterns and structures in the objective function, constraint matrices, or instances (Punnen et al. 2017, Bettiol et al. 2022). In this paper, we employ such a technique to exploit the structure of the quadratic cost matrix $\boldsymbol{q}$. Our methodology relies on splitting the quadratic cost $\boldsymbol{q}$ into (i) in-star interactions in which the quadratic costs between adjacent edges are investigated and (ii) out-of-star interactions corresponding to the quadratic costs between nonadjacent edges. This leads to a new reformulation of BQP with exponentially many variables, each associated with a star in G. Inspired by the star-reformulation of Pereira et al. (2013) for the adjacent (QMSTP) spanning tree problem, we develop a column generation (CG) to derive bounds for some classes of BQPs with different levels of in-star and out-of-star quadratic costs.

1.1. Literature Review

One of the most natural ways to solve the BQP problems with an exact method is to linearize the quadratic terms of the model and solve the resulting mixed-integer linear programming (MILP) using state-of-the-art solvers. The standard linearization technique (SLT) is one of the most well-known linearizations in the literature of BQPs (Glover and Woolsey 1974). However, two main concerns appear when dealing with the MILP reformulation: the increased size of the problem (in terms of variables and constraints) and the quality of the obtained dual bounds. There have been many attempts to deal with these concerns in the literature. Adams and Sherali (1990), Adams and Forrester (2005), and Sherali and Smith (2007) provide different reduced-size MILP reformulations of the BQP, while Liberti (2007) introduces a compact linearization approach for a general class of binary quadratic problems subject to assignment constraints. This approach is then revised in Mallach (2018) by proposing two new necessary and sufficient conditions to achieve consistent linearization for this class of problems. In another study, Jünger and Mallach (2021) investigate the solution methods proposed in the literature for unconstrained BQPs based on linear programming (LP). They provide some enhancements to the algorithm of the central separation problem arising in solving these problems. Following this study, Charfreitag et al. (2022) present a solver called McSparse, based on a branch-and-cut algorithm to obtain exact solutions for sparse unconstrained binary quadratic programming problems. Furthermore, Hahn et al. (2012), Sherali and Adams (2013), and Rostami and Malucelli (2015) develop the reformulation-linearization technique (RLT), which generally provides stronger MILP reformulations. Recently, Mallach (2023) investigated the effectiveness of the inductive linearization technique for some applications of BQPs. Mallach (2023) represents that although this linearization technique is more compact in terms of constraints, the continuous relaxation is at least as tight as the standard linearization technique.

Semidefinite programming (SDP), quadratic reformulation, and cutting-plane methods are alternative approaches used to generate strong relaxations of BQP. In SDP, which is considered an extension of MILP reformulations, nonnegativity constraints are replaced by positive semidefiniteness constraints (Helmberg et al. 2000, Lemaréchal and Oustry 2001). In quadratic reformulations, one must alter the objective function of a BQP problem and transform it into an equivalent convex/nonconvex BQP problem to generate tighter dual bounds (Billionnet et al. 2009, Rostami et al. 2023). The use of valid inequalities, which are generated and added in a cutting-plane fashion, is another approach commonly adopted in the literature (Fischer 2014).

Another relevant approach to obtain a stronger reformulation for the BQP is to use decomposition techniques. Variants of decompositions such as Lagrangian decomposition, graph partitioning, and CG methods are employed to explore the bounds of unconstrained BQP problems (Mauri and Lorena 2011, 2012). For constrained BQP problems, Chen et al. (2018) represent bounds for the BQP using a Lagrangian-based heuristic method. Some examples of using decompositions to tackle the problem of investigating bounds for some specific BQP problems are observed for the quadratic knapsack problem and the minimum spanning tree problem (Billionnet and Soutif 2004, Pereira and da Cunha 2020).

There are a few papers in the literature reformulating a BQP model for a specific application into a MILP with an exponential number of variables, which is solved by CG. Aloise et al. (2010) reformulate the mixed 0-1 quadratic programming model of the modularity maximization problem and solve the reformulation using a stabilized CG. In a related study, a CG heuristic is used in a districting problem to produce the best territories for the purpose of financial product pricing (De Fréminville et al. 2015). Rostami et al. (2016) propose a lower bounding procedure for the

asymmetric quadratic traveling salesman problem. They reformulate the problem as a MILP with an exponential number of cycles as variables and solve the relaxation using a CG. Recently, Yarkony et al. (2020) developed an extended MILP formulation for correlation clustering. They consider solving correlation clustering for several computer vision applications through CG, Benders decomposition, and dynamic programming (DP).

The works most closely related to our study are the papers devoted to the reformulation of the adjacent-only quadratic minimum spanning tree problem (AQMSTP) to a MILP and solution algorithms applied to solve the reformulation. More specifically, a star-reformulation is introduced in Pereira et al. (2013) based on the concept of stars in graph theory to reformulate the AQMSTP by a stronger LP. Then, an algorithm based on dynamic column and row generation is proposed to determine the dual bounds of the problem. In another study, Pereira et al. (2015) present a branch-and-cut-and-price algorithm based on this reformulation of the AQMSTP and a branch-and-cut algorithm based on projecting out the decision variables of this model. There are also other reformulations and algorithms for the AQMSTP in the literature based on the proposed star-based model (e.g., Pereira and da Cunha 2018, 2020).

In recent years, identifying patterns and structures inherent in large-scale optimization problems and using them to efficiently tackle these real-life problems have drawn researchers' attention. Exploring the constraint matrix structures (Punnen et al. 2019), finding important patterns in a specific problem class (Khaniyev 2018), and revealing the structure of a data instance related to a large-scale problem (Khaniyev et al. 2020) are among the notable techniques in this area. One of the structures explored in the MILP literature is the singly bordered block diagonal (BBD) structure. Exploiting this structure in the constraint matrices of a MILP leads to Dantzig-Wolfe (DW) decomposition, Lagrangian relaxation, and branch-and-price (Bergner et al. 2015, Khaniyev et al. 2018). Bergner et al. (2015) provide a computational proof-of-concept to show that the DW reformulation can be automated and applied to all MILPs by exploiting and rearranging the structure of the constraint's matrix in various ways. They suggest a score to measure the quality of each decomposition and identify the most useful one for the DW reformulation of a mixed-integer programming (MIP). To our knowledge, there are few methodological studies concerning structures in BQP problems. Bettiol et al. (2022) tackle BQP from the CG perspective and construct an approach to study the structure of block-decomposable problems in BQP. They present two types of relaxations that acquire strong lower bounds (LBs) for general BQP and block-decomposable BQP specifically. The relaxations are based on DW reformulation, although CG is used as their solution method. In addition, some studies explore the linearizability of the quadratic cost matrices. Punnen et al. (2017) investigate the structure of quadratic cost matrices to propose necessary and sufficient conditions for linearizability of the quadratic traveling salesman problem. In another related work, Hu and Sotirov (2021) present a linearization-based lower bounding scheme applicable to several BQP problems using a certificate for a quadratic function to be nonnegative on the feasible set.

To the best of our knowledge, the idea of exploring the effectiveness of the star-reformulation on more general cases rather than the AQMSTP has not been investigated in the literature. Our major focus here is to explore the generalizability and usability of this reformulation by splitting the quadratic cost to model and solve more general problems in BQP.

1.2. Main Contributions

Our main contributions are summarized as follows:

- We investigate the idea of the star-reformulation proposed for the AQMSTP in the literature on the adjacent-only BQP problems. Moreover, we show how to exploit such a structure to split the general BQP problems' quadratic costs into adjacent and nonadjacent components.
- We develop a CG for each specific reformulation to derive valid dual bounds.
- To demonstrate the potential of the star-reformulation, the cost-splitting framework, and the solution methodology, we consider three BQP problems (quadratic semi-assignment, adjacent-only quadratic semi-assignment, and multiple object tracking (MOT)) whose reformulations lead to different master problems and pricing subproblems, that is, (i) quadratic master problem and unconstrained BQP pricing subproblem, (ii) linear programming master problem and unconstrained BQP pricing subproblem, and (iii) linear programming master problem and constrained BQP pricing subproblem.
- We perform extensive computational experiments to evaluate the proposed reformulations and CG algorithms on the adjacent-only and general BQP problems. The special quadratic matrix structure of adjacent-only BQPs and BQP problems with sparse quadratic costs derive the most significant benefit from the star-reformulation and the proposed splitting framework.

The rest of the paper is organized as follows. In Section 2, we present the idea of star-reformulation and introduce the adjacent-only BQP class of problems. Section 3 corresponds to the proposed cost-splitting framework, and then we explain a column generation to solve the reformulation. Section 4 is dedicated to presenting and reformulating one BQP example to evaluate the cost-splitting reformulation and two adjacent-only BQP problems, the adjacent-only

quadratic semi-assignment problem (AQSAP) and MOT, as our illustrative examples to generalize star-reformulation. Finally, in Section 5, we explain our test sets and display the computational results from the introduced problems, and in Section 6, we wrap up the paper and provide some future possible directions for research.

2. Star-Reformulation for Adjacent-Only BQP Problems

In many real-life applications modeled as BQP on graphs, the quadratic costs appear only between adjacent edges. As such, interaction costs are zero for pairs of edges that do not share a common endpoint. Some notable examples include the adjacent-only quadratic minimum spanning tree problem (Pereira et al. 2013, Pereira and da Cunha 2020), the quadratic traveling salesman problem (Fischer 2014, Punnen et al. 2017), the adjacent quadratic assignment problem (Fischer et al. 2009), the adjacent quadratic shortest path problem (Rostami et al. 2015, Hu and Sotirov 2018), and variants of the correlation clustering and the modularity maximization problems (Bonizzoni et al. 2008, Aloise et al. 2010, Yarkony et al. 2020).

To exploit such a unique structure in developing a solution strategy, we follow Pereira et al. (2013) to reformulate the adjacent-only BQP to a new model with exponentially many variables based on the concept of stars in a graph. To this end, we first provide some notations to introduce the concept in Section 2.1 and then present the star-based reformulation in Section 2.2. For further information on the star-reformulation, we refer the readers to Pereira et al. (2013), where the idea is introduced for the adjacent-only QMSTP.

2.1. Definitions

Consider the BQP problem on graph $G = (V, E)$ in (1). Without loss of generality, we assume that $V = N \cup H$ where N and H are disjoint sets. We also take into account the possibility of N being an empty set. Thus, in this situation, the set V is equal to the set H. For each $v \in H$, we define $\delta(v) \subseteq E$ as the set of edges incident to node v and let $A = \cup_{v \in H} \delta(v)$ be the set of all edges with exactly one endpoint in H, considering exceptionally, when the set N is empty, both endpoints of all edges of the related graph are thereby included in H. So, if $N \neq \emptyset$, A is the set of edges with exactly one endpoint in H, and if $N = \emptyset$, then $A = E$.

Two distinct edges of A, say $e = \{i, j\}$ and $f = \{k, \ell\}$, are adjacent if they share a common endpoint v in H, that is, if $\{i, j\} \cap \{k, \ell\} = v \in H$. We denote by $\mathcal{A}$ the set of distinct pairs of adjacent edges in A:

$$\mathcal{A} = \{(e, f) \in A \times A \ : e = \{i, j\}, f = \{k, \ell\}, \ \{i, j\} \cap \{k, \ell\} = v \in H\} \tag{2}$$

The graphs indicated in Figure 1 illustrate the concept of adjacent edges by defining sets A and $\mathcal{A}$.

These definitions will lead to the following definition for the star-shaped subgraph s. That is, for each $v \in H$ we define a star s centered at node v as any subset of $\delta(v)$ whose elements are in the set A, including an empty set, and let S^v be the set of all stars centered at node v. The mathematical representation of the aforementioned definitions can be expressed as

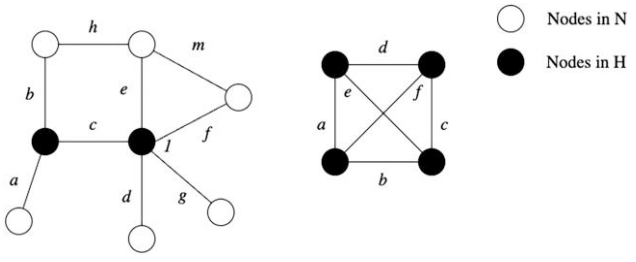
$$S^v = \{s : s \subseteq (\delta(v) \cap A)\}.$$

Therefore, $S = \cup_{v \in H} S^v$ includes all the possible stars centered at nodes $v \in H$ in the graph. As an example, in the left graph of Figure 1, $s = \{e, f, g, d\}$ is the largest possible star centered at node 1. Note that as both endpoints of the edge c are in the set H, this edge is not included in any possible star of this graph.

2.2. Star-Reformulation

In the adjacent-only cases, interaction costs are zero for pairs of edges that do not share a common endpoint in the defined subset H. Because of Equation (2), in this BQP class, the quadratic cost $\boldsymbol{q}$ for the pairs of edges (e, f) that are not

Figure 1. Two Examples of Graphs to Demonstrate the Concepts of Set A, Adjacency Set $\mathcal{A}$, and Star s



Notes. Graph on the left: $A = \{a, b, d, e, f, g\}$, $\mathcal{A} = \{(a, b), (d, e), (d, f), (d, g), (e, f), (e, g), (f, g)\}$. Graph on the right: $A = \{a, b, c, d, e, f\}$, $\mathcal{A} = \{(a, b), (a, d), (a, e), (a, f), (b, c), (b, e), (b, f), (c, d), (c, e), (c, f), (d, e), (d, f)\}$. Node 1 is specified to motivate a star center candidate.

covered by the set $\mathcal{A}$ is zero. Based on the represented concepts and the fact that each star $s \in S$ consists of a subset of pairs in $\mathcal{A}$, we can reformulate an adjacent-only BQP problem in terms of stars. For each $s \in S$ let $C_s = \sum_{e \in s} c_e + \sum_{e,f \in s} q_{ef}$ represent the total linear and quadratic cost of star s. We define a new binary decision variable λ_s for each star $s \in S$ to indicate if the corresponding star s is included in the solution of the BQP problem or not. We also denote $\mathcal{F}$ to depict the feasible space of the problem. By preserving the definition of variable $\boldsymbol{x} \in X$, we provide the following star-based reformulation:

$$\min \quad \sum_{s \in S} C_s \lambda_s + \sum_{e \in E \setminus A} c_e x_e \tag{3}$$

$$\text{s.t.} \quad (\boldsymbol{x}, \boldsymbol{\lambda}) \in \mathcal{F}(\boldsymbol{x}, \boldsymbol{\lambda}) \tag{4}$$

$$\boldsymbol{x} \in \{0,1\}^m \tag{5}$$

$$\boldsymbol{\lambda} \in \{0,1\}^{|S|} \tag{6}$$

The objective function of the reformulation minimizes the total cost of the problem and consists of two different parts. The first part corresponds to the cost of star s, including the linear costs of the edges inside the star and the interaction between adjacent edges of that star. The second term of the objective function reflects the linear cost of the edges which are not incorporated in any possible star. Constraint (4) links the feasible region of the problem to the stars by coupling the original variables $\boldsymbol{x}$ and new variables $\boldsymbol{\lambda}$. It can also include the constraints which are only related to the variables $\boldsymbol{\lambda}$ and the constraints which are only associated with variables $\boldsymbol{x}$. We assume, without loss of generality, that such linking constraints can always be found, because for each $v \in H$ and each $e \in \delta(v)$, there exist parameters $b_{es} \in [0,1]$ such that $x_e = \sum_{s \in S^v} b_{es} \lambda_s$ (e.g., see Section 4). Because some of the constraints in the original model can be included in the definition of the star in this reformulated counterpart, the $\boldsymbol{x}$-only constraints can be considered as a subset of X in (1).

To elaborate on the star-reformulation, consider the right graph of Figure 1. In this graph, a given set of star centers (H) potentially comprises the whole node set of the graph (V) (i.e., $H = V$). Intuitively, for BQP problems with this property, set A contains all the graph's edges. To illustrate this, consider the AQMSTP (Assad and Xu 1992), in which the set H is equal to all nodes of the graph ($N = \emptyset$). Thus, in the reformulated model of the AQMST of Pereira et al. (2013), the objective function consists of only a linear term to represent the cost of stars. Nonetheless, in the left graph of Figure 1, some of the nodes are not considered as possible star centers ($N \neq \emptyset$), so we keep the second term of the reformulation's objective function to demonstrate the linear cost of the edges which are not included in any possible star. In this study, we employ the star-reformulation to model two applications; more details on modeling and solution methodology are presented in Section 4.

3. Cost-Splitting and Star-Reformulation for General BQP Problems

In order to tackle the complexity of the general BQP problem of (1) using decomposition and inspired by the idea of the star-reformulation of Section 2, we propose a cost-splitting framework to reformulate BQP problems. Our methodology relies on the concept of star structures to partition the objective function of (1) at any feasible solution $\overline{\boldsymbol{x}}$ into four parts: in-star linear costs, out-of-star linear costs, in-star quadratic costs, and out-of-star quadratic costs. More precisely, given a feasible solution $\overline{\boldsymbol{x}} = (\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_{m_1}, \ldots, \overline{x}_m) \in X \subseteq \{0,1\}^m$, we can rewrite it as $\overline{\boldsymbol{x}} = (\overline{\boldsymbol{x}}^1, \overline{\boldsymbol{x}}^2)$ with $\overline{\boldsymbol{x}}^1 \in \{0,1\}^{m_1}$ and $\overline{\boldsymbol{x}}^2 \in \{0,1\}^{m-m_1}$ and where $\overline{\boldsymbol{x}}^1$ is related to feasible solutions of edges in set A and $\overline{\boldsymbol{x}}^2$ stands for feasible solutions of edges in $E \setminus A$. Therefore, the objective function of (1) at $\overline{\boldsymbol{x}}$ can be written as

$$\sum_{e \in A} c_e \overline{x}_e + \sum_{e \in E \setminus A} c_e \overline{x}_e + \sum_{(e,f) \in \mathcal{A}} q_{ef} \overline{x}_e \overline{x}_f + \sum_{(e,f) \in \mathcal{E} \setminus \mathcal{A}} q_{ef} \overline{x}_e \overline{x}_f.$$

As presented in Section 2, the first and third terms of this formula together depict the costs of inside stars, whereas the last term as a quadratic cost is associated with the interaction between pairs of nonadjacent edges. Therefore, the cost-splitting reformulation of the BQP problem of (1) is proposed below where the constraints are maintained as in (4)–(6):

$$\min \quad \sum_{s \in S} C_s \lambda_s + \sum_{e \in E \setminus A} c_e x_e + \sum_{(e,f) \in \mathcal{E} \setminus \mathcal{A}} q_{ef} x_e x_f \tag{7}$$

$$\text{s.t.} \quad (4)\text{–}(6). \tag{8}$$

Although modeling BQP problems based on separating adjacent edge interactions and nonadjacent edge interactions is the principal notion behind the proposed model, each term of the objective function can potentially be eliminated based on the specific problem structure in different applications. Nonetheless, we keep the first term of the objective function as the basis of our reformulation. As an example, in the quadratic minimum spanning tree (QMST) problem

(Assad and Xu 1992), because the quadratic cost comprises the interactions between all pairs of edges (adjacent and nonadjacent) and the set of possible star centers consists of the complete node set V, in the objective function of the cost-splitting reformulated model of the QMST problem, we incorporate a linear term to represent the cost of stars and a quadratic term for interactions between nonadjacent edges. Another notable example is the uncapacitated single allocation p-Hub median problem (USApHMP) (O'Kelly 1987, Meier et al. 2016), which can be formulated as a BQP problem. This problem can be reformulated using our star-based model, consisting of linear costs inside the possible stars and quadratic interactions between stars with different centers.

3.1. Column Generation Approach

The use of the star representation in the proposed reformulation of the BQP in (7)–(8) results in an exponential number of variables $\lambda_s, s \in S$. Pereira et al. (2013) suggest a column generation for the star-reformulation of the AQMST problem. Inspired by this idea, CG is applied to deal with the proposed cost-splitting of (7)–(8).

CG is an efficient iterative algorithm for providing dual bounds for problems with an exponential number of variables (columns) (Dantzig and Wolfe 1960). In each iteration of the algorithm, it solves one restricted master problem (RMP), which is the problem restricted to a small subset of the variables, and one or several pricing subproblems. Using the dual information of the RMP, the pricing subproblem is solved to verify the optimality of the master problem and the CG algorithm stops if the optimality condition is satisfied. Otherwise, one or more new variables determined by the subproblem will be added to the RMP and the updated RMP will be solved in the new iteration.

Because the generic reformulation (7)–(8) is a nonconvex quadratic problem, the standard CG procedure cannot be directly applied. If the model is convex, meaning that the matrix $\boldsymbol{q}$ is positive semidefinite, then the underlying nature of that problem brings out either a convex quadratic subproblem or a convex quadratic master problem. In the case of a convex RMP, the primal and dual solutions of the RMP can be obtained by state-of-the-art solvers. However, in our proposed framework, we do not restrict the definition of $\boldsymbol{q}$ to positive semidefinite matrices. Therefore, one has to deal with the quadratic term of the objective function. Different types of convexification (Billionnet et al. 2009), linearization, semidefinite programming relaxation, and BQP relaxation for block-decomposable problems (Bettiol et al. 2022) are alternative options for handling the BQP master problem. Nevertheless, some of these methods require additional complex constraints and variables which may adversely affect their performance. Our approach is based on the idea of converting the quadratic cost to linear cost by separating the in-star and out-of-star costs and then replacing each of them by a linear term. Here, we follow linearization techniques to transform the quadratic objective function into an equivalent MILP. We define y_{ef} as the linearized variable to replace the quadratic terms $x_e x_f$, $(e,f) \in \mathcal{E} \setminus \mathcal{A}$, impose a set of linking constraints $\mathcal{P}(\boldsymbol{x}, \boldsymbol{y})$ to guarantee $y_{ef} = x_e x_f$, and consider $\overline{S} \subseteq S$ as a subset of possible stars, so we obtain the following RMP for the LP relaxation of the problem:

$$\min \quad \sum_{s \in \overline{S}} C_s \lambda_s + \sum_{e \in E \setminus A} c_e x_e + \sum_{(e,f) \in \mathcal{E} \setminus \mathcal{A}} q_{ef} y_{ef} \tag{9}$$

$$\text{s.t.} \quad (\boldsymbol{x}, \boldsymbol{\lambda}) \in \overline{\mathcal{F}}(\boldsymbol{x}, \boldsymbol{\lambda}) \tag{10}$$

$$(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{P}(\boldsymbol{x}, \boldsymbol{y}) \tag{11}$$

$$\boldsymbol{y} \in \mathbb{R}_+^{|(e,f) \in \mathcal{E} \setminus \mathcal{A}|} \tag{12}$$

$$\boldsymbol{\lambda} \in [0,1]^{|\overline{S}|} \tag{13}$$

where $\overline{\mathcal{F}}(\boldsymbol{x}, \boldsymbol{\lambda})$ is a subset of $\mathcal{F}(\boldsymbol{x}, \boldsymbol{\lambda})$ restricted to $\overline{S}$.

In each iteration, we add a subset of columns $s \in S \setminus \overline{S}$ which potentially improves the objective function of (9). To this end, we solve an auxiliary problem to find the most negative reduced cost column to add to the master problem. Thus, a column entering the basis can be found by computing the minimum reduced cost star with respect to the quadratic and linear costs of the edges inside the star.

According to the definition of the cost of stars C_s, the pricing subproblem can be either a linear problem or a BQP problem. In a simple setting, this subproblem can take the form of an unconstrained BQP (UBQP) problem. Nevertheless, it is possible that one must explicitly incorporate constraints in the binary quadratic subproblem in some specific applications. Given that the essence of solving a BQP problem exactly is NP-hard, intuitively adding columns with negative reduced cost without solving the subproblem to optimality can be a promising alternative when applying a CG algorithm. However, to provide a/some valid dual bound(s), we need to solve the subproblem to optimality (Aloise et al. 2010, De Fréminville et al. 2015). Specifically, in the case of a UBQP subproblem, several solution approaches based on greedy and heuristic methods are proposed to solve this problem (Kochenberger et al. 2014). Even in the case where the pricing is a constrained BQP problem and obtaining an exact solution is necessary, the size

of the problem is much smaller than the compact formulation in Section 1, meaning the problems are relatively easy to solve. In addition, when the subproblem is a constrained BQP problem, we have better options in terms of linearization techniques such as RLT to solve it exactly. In the following section, we use examples to demonstrate how to deal with each of these cases.

Although the model presented in (7)–(8) is a generic reformulation of the BQP, special structure assumptions for the quadratic matrix $\boldsymbol{q}$ to have only adjacent quadratic costs offer promising properties for solving this type of problems. The third term of the objective function in (9) is therefore eliminated, and the stars interact with each other only through linear costs. The new formulation for this particular class is an integer linear problem, and consequently the constraints (11) and (12) are removed, leading to a model which is more tractable to solve than the generic model. Note that the quadratic interaction between two adjacent edges may depend on their common endpoint, or may be independent of it, resulting in different sparsity levels of the matrix $\boldsymbol{q}$.

4. Illustrative Examples

The objective of this section is to demonstrate how the star-reformulation, the cost-splitting framework, and the solution strategy described in Sections 2 and 3 can be applied to different BQP problems. To this end, we consider three BQP problems whose formulations lead to different master problems and pricing subproblems, described as follows:

- Quadratic master problem and unconstrained BQP pricing subproblem
- Linear programming master problem and unconstrained BQP pricing subproblem
- Linear programming master problem and constrained BQP pricing subproblem

We consider the quadratic semi-assignment problem (QSAP) as an example to demonstrate the performance of the cost-splitting idea in addition to two problems from the adjacent-only class of the BQP, the AQSAP and the MOT problem, to outline the computational advantages of the star-reformulation in this class. In the following subsections, we provide a brief description and literature review for each problem, as well as the compact BQP models. Then, we describe how to reformulate them as (7)–(8) and obtain dual bounds using our CG solution method.

4.1. QSAP

In this problem, we are given a set of clients $N = \{1, \dots, n\}$ and a set of servers $H = \{1, \dots, h\}$. Suppose there is a linear cost $c_e, e = \{i, j\}$, associated with the assignment of client $i \in N$ to server $j \in H$, and there is a quadratic cost $q_{ef}, e = \{i, j\}, f = \{k, l\}$, associated with the assignment of client $i \in N$ to server $j \in H$ and client $k \in N$ to server $l \in H$ simultaneously. We transfer this problem to the previously defined graph G, with the node set V consisting of n clients and h servers and the edge set $E = \{(i, j) \mid i \in N, j \in H\}$. By recalling the concepts of Section 3, we define the binary decision variable as x_e. Hence, denoting by variable x_e equals one if the edge e is chosen (the client i is assigned to server j), and zero otherwise, the BQP problem of semi-assignment on graphs is formulated as follows:

$$\min \quad \sum_{e \in A} c_e x_e + \sum_{(e,f) \in \mathcal{E}} q_{ef} x_e x_f \tag{14}$$

$$\text{s.t.} \quad \sum_{e \in \delta(i)} x_e = 1 \qquad \forall i \in N \tag{15}$$

$$x_e \in \{0, 1\} \qquad \forall e \in E. \tag{16}$$

The QSAP has a variety of applications in the area of scheduling (Stone 1977, Chrétienne 1989) and partitioning (Hansen and Lih 1992). The hub network design problem is also considered as a special case of the QSAP (Saito et al. 2009). The problem is known to be NP-hard, and solving it even for small-size instances is very time-consuming (Sahni and Gonzalez 1976, Magirou and Milis 1989, Malucelli 1996). Using RLT is a common approach in the literature to solve the QSAP, and there are also some studies on polynomial algorithms, heuristics, and lower bounding methods for special cases of the QSAP. We refer the reader to Saito et al. (2009) and the references therein for more details.

4.1.1. Reformulation. We reformulate the QSAP as the general model (7)–(8). To this end, without loss of generality, we assume that every server j is a center of a star-shaped subgraph s. So, the binary variable λ_s corresponds to selecting this star. We define parameter $B_{js} \in \{0, 1\}$ to indicate if server j is the center of star s or not, the parameters $D_{is} \in \{0, 1\}$ to identify if client i is included in star s, and $D_{es} \in \{0, 1\}$ to denote whether edge e belongs to star s or not. The out-of-star interactions in the QSAP result in a quadratic reformulation, so based on the notation given above, and according to the formulation (9)–(13), the linearized RMP can be expressed as follows:

$$[\text{RMP-QSAP}]: \quad \min \sum_{s \in \bar{S}} C_s \lambda_s + \sum_{(e,f) \in \mathcal{E} \setminus \mathcal{A}} q_{ef} y_{ef} \tag{17}$$

$$\text{s.t.} \quad \sum_{s\in\overline{S}} B_{js}\lambda_s \le 1 \qquad \forall j \in H \tag{18}$$

$$\sum_{s\in\overline{S}} D_{is}\lambda_s = 1 \qquad \forall i \in N \tag{19}$$

$$\sum_{s\in\overline{S}} D_{es}\lambda_s = x_e \qquad \forall e \in A \tag{20}$$

$$(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{P}(\boldsymbol{x},\boldsymbol{y}) \tag{21}$$

$$0 \le x_e \le 1 \qquad \forall e \in A \tag{22}$$

$$\boldsymbol{y} \in \mathbb{R}_+^{|(e,f)\in\mathcal{E}\setminus\mathcal{A}|} \tag{23}$$

$$\boldsymbol{\lambda} \in [0,1]^{|\overline{S}|} \tag{24}$$

where y_{ef} are the linearization variables. Constraints (18) impose that at most one star can be chosen among all the stars centered at j. Constraints (19) are the set partitioning linking constraints, which impose that each client must be included in exactly one star. Constraints (20) are the linking constraints and enforce that if an edge is selected in an optimal solution, then it has to be included in only one selected star.

4.1.2. Column Generation. Starting from a subset of stars ($\{s \mid \lambda_s = 1\}$) which are feasible with respect to the constraints of the reformulation as initial columns, the algorithm solves the [RMP-QSAP] restricted to the current set of stars in each iteration. The corresponding dual solutions construct one pricing subproblem for each server $j \in H$, which aims to find a star related to j with the minimum reduced costs. In the next iteration of the algorithm, these stars are added to [RMP-QSAP] as the new columns to possibly improve the objective value of the master problem. The algorithm terminates when there are no more columns with a negative reduced cost to be added.

Let π_j, ρ_i, and γ_{ij} be the dual solutions corresponding to constraints (18)–(20), respectively. For each server j we consider a set of incident edges to that server. Therefore, we can use them to rewrite the dual solutions $\gamma_{i,j}$ as γ_e, where $e = \{i,j\}$. In addition, ρ_i is not dependent on j, and is fixed for all star centers, so we use $\rho_e, e = \{i,j\}$ instead of ρ_i in our formulation to keep it consistent with the rest of the terms.

The star with minimum reduced cost can be found by solving the following pricing subproblems on graphs, one for each server j:

$$\min \quad \sum_{e\in\delta(j)} (c_e - \rho_e - \gamma_e) z_e + \sum_{e,f\in\delta(j)} q_{ef} z_e z_f - \pi_j \tag{25}$$

$$\text{s.t.} \quad z_e \in \{0,1\} \qquad \forall e \in \delta(j) \tag{26}$$

where binary decision variable $z_e, e = \{i,j\}$ indicates if client i is part of the star centered at server j.

4.2. AQSAP

In this section, we consider a special class of the QSAP in which the quadratic costs are restricted to the adjacent edges only. Consider the QSAP of assigning n clients to h servers, this time in a distributed processing system. If client $i \in N$ is assigned to server $j \in H$, the required processing time c_{ij} is computed based on the processing speed of the server and the client's demand. In this type of problem, where multiple clients are assigned to the same server, there is no predefined priority and the order of processing the requirements is unknown. In this situation, every client $i \in N$ aims to minimize the worst-case completion time. Hence, the completion time of client i is set to $CT_i = \sum_{j\in H} x_{ij} \sum_{k\in N} c_{kj} x_{ij} x_{kj}$, where the binary variable $x_{ij} = 1$ indicates assigning client i to server j (Drwal 2014). The goal in this problem is obtaining an assignment to minimize the total completion time for all clients, $\sum_{i\in N} CT_i$. According to the definitions presented in Section 4.1 and considering edge $e = \{i,j\}$ and $f = \{k,j\}$, the objective function of this problem on the graph G is

$$\min \quad \sum_{e\in A} c_e x_e + \sum_{(e,f)\in\mathcal{A}} q_{ef} x_e x_f \tag{27}$$

where the constraints of the problem are the same as (15) and (16) and the equation below holds:

$$q_{ef} = c_e + c_f. \qquad \forall (e,f) \in \mathcal{A} \tag{28}$$

This formulation is valid for all assignment problems in which multiple clients compete for a single machine and each assigned client has to undergo the completion time of the machine. This problem includes a large class of the QSAP, although the structure of its quadratic matrix narrows it down to the class of problems described in Section 2. We

denote this problem by AQSAP in the rest of the paper. Note that the explanations provided here are based on an application of AQSAP for clarity. Nevertheless, AQSAP is a general problem and can be applied in various applications such as specialized variants of scheduling and selfish resource allocation on the internet. We observe two fundamental properties of the current problem: (i) similar to the QSAP, the linear costs for the edges which are not covered by stars are zero, and (ii) the nonadjacent edges do not interact with each other. Therefore, there are no out-of-star interactions between edges in the AQSAP, which in turn leads to the following reformulation:

$$\min \quad \sum_{s\in S} C_s\lambda_s \tag{29}$$

$$\text{s.t.} \quad \sum_{s\in S} B_{js}\lambda_s \le 1 \qquad \forall j \in H \tag{30}$$

$$\sum_{s\in S} D_{is}\lambda_s = 1 \qquad \forall i \in N \tag{31}$$

$$\boldsymbol{\lambda} \in [0,1]^{|S|} \tag{32}$$

In this case, the reformulation is linear and CG is directly applicable. In order to solve the reformulation with CG, the pricing subproblem on graphs for every star center j can be written as

$$\min \quad \sum_{e\in\delta(j)} (c_e - \rho_e)z_e + \sum_{e,f\in\delta(j)} q_{ef}z_ez_f - \pi_j \tag{33}$$

$$\text{s.t.} \quad z_e \in \{0,1\} \qquad \forall e \in \delta(j) \tag{34}$$

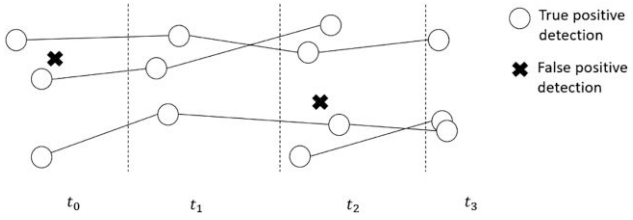
where $\pi_j, j \in H$, and $\rho_i, i \in N$ are the optimal duals associated with constraints (30) and (31), respectively, and we transfer them on the edges incident to j. Similar to the QSAP, here we end up with a UBQP pricing subproblem. Implementation details of the CG approach, and extensive computational experiments to find dual bounds, are provided in Section 5.

4.3. Multiple Object Tracking

MOT and, more specifically, multiple person tracking are a well-known application in computer vision that aims to track multiple objects (people) in a sequence of video frames. MOT is associated with a variety of applications like self-driving cars, human-computer interaction, security and video surveillance, sports analysis, some games like Microsoft Kinect, traffic analysis, etc. (Emami et al. 2018, Shen et al. 2018). Despite recent developments in this area, it is still a very challenging task because of occlusion and scene cluttering. As a result of the advancement of object detection technologies, detection-based methods are the most dominant techniques in MOT (Tang et al. 2017, Shen et al. 2018). MOT consists of three main components: (i) detecting the objects, in which a person detector is utilized for each individual frame to find the potential locations of all the people; (ii) affinity, or score estimation, which demonstrates how likely detections are related to a single identity; and (iii) data association, in which these hypotheses are linked across the frames based on the estimated scores to form tracks (Henschel et al. 2018).

In general, whereas object detection and score determination are deep learning tasks, data association is a combinatorial optimization problem. Once the detections and their unary and pair-wise scores are computed, they are given as inputs to the data association problem to generate the associated tracks. In the context of MOT, a sequence of frames (from t_0 to t_3) of a scene containing a few objects (such as people) is depicted in Figure 2. These potential objects are identified as a priori through the use of a machine-learning-based object detection method which provides a list of potential objects and the likelihood scores associated with them. This list and the detection scores are then used as the inputs in the optimization model. The detected objects are represented by circles and cross marks in this figure where the detection scores produced by the detection algorithm are included as a part of the linear and quadratic costs in the

Figure 2. Graphical Representation of a Solution of the Data Association in MOT Applied to Detections of Four Video Frames



MOT model. This figure represents the results of the data association step. In this step, an optimization model is solved to determine the set of detections that belong to the same object across multiple frames based on the cost parameters. The figure shows that, after solving the optimization model, the solution indicates four tracks (people) which are depicted by connecting lines in the figure and two false detections, represented by cross marks. As shown in Figure 2, two of the objects are captured in all four frames while one of the objects moves through frames t_0 to t_2 and the other object moves only from frame t_2 to the last frame.

Data association algorithms can be categorized as online or offline algorithms (Emami et al. 2018). Here we consider the offline data association case.

Essentially, a data association problem can be modeled with respect to a graph $G = (V, E)$. A detection $i \in N$ is represented by a node in this graph. We consider another set of nodes $H = \{1, 2, \ldots, h\}$, which are dummy nodes related to the tracks (target people). h is an upper bound (UB) on the number of target people in the video, which is predefined as an input. More precisely, we define the graph where the vertex set V consists of all detections, that is, potential bounding boxes of potential candidates of people N in a video sequence and all possible tracks (target people) H. Similar to Section 2, consider A as a subset of edges E which are incident in a node in H. Therefore, edge $e = \{i, j\} \in A$ denotes a possible linking of a detection to a track (person).

Given $T = \{1, 2, \ldots, \mathcal{T}\}$ as the set of all frames, each detection i belongs to a frame $t \in T$. We introduce two other subsets of edges based on our definitions in Section 2.1: $\delta(i) \subseteq A$ is a subset of edges in A incident to node i and $\delta^t(i)$ is a subset of $\delta(i)$ when the edges stem from the frame t.

Each edge $e = \{i, j\} \in A$ has a cost $c_e \in \mathbb{R}$ defined via a logit function and reflects the likelihood of detection $i \in N$ being a correct detection. This cost is called unary cost in the computer vision literature (Henschel et al. 2018). Unary cost is fixed for each detection i and is not dependent on tracks. For every pair of edges (e, f) which are incident in a node in H, a pair-wise cost $q_{ef} \in \mathbb{R}^{m \times m}$ is to be paid. Note that the interaction between edges e and f is nonzero if, and only if, the detections $i \in N$ and $k \in N$ are assigned to a distinct track. Pair-wise cost identifies how likely it is that two detections belong to the same person. The probabilities which determine the costs are inferred based on detection scores. There are several ways to estimate these score terms from the geometry features, color histogram, appearance, and other features related to the image data. Considering p as this probability, the quadratic cost is computed as $log(1 - p)/p$, so depending on the unary and pair-wise probabilities, the costs can be negative, positive, or zero, resulting in nonconvexity of the problem (Dehghan and Shah 2017, Henschel et al. 2018).

Similar to the model presented in Henschel et al. (2018), the MOT data association problem to minimize the total cost of labeling is expressed as

$$\min \quad \sum_{e \in A} c_e x_e + \sum_{(e,f) \in \mathcal{A}} q_{ef} x_e x_f \tag{35}$$

$$\text{s.t.} \quad \sum_{e \in \delta(i)} x_e \leq 1 \qquad \forall i \in N \tag{36}$$

$$\sum_{e \in \delta^t(j)} x_e \leq 1 \qquad \forall j \in H, \quad \forall t \in T \tag{37}$$

$$x_e \in \{0, 1\} \qquad \forall e \in A \tag{38}$$

where constraints (36) are needed to mandate that more than one track assignment is not possible for every detection i. Constraints (37) restrict the model to select at most one detection associated with each track inside every frame.

Although the BQP model from Henschel et al. (2018) contains only constraints (36), different types of potential side constraints can be added to this primary model. These constraints can be formulated based on prior knowledge or hypotheses associated with the scenes that video frames come from, the types of cameras taping the frames, or the features of the object detector as well as other constraints based on the prior knowledge with respect to the objects in the scenes (Assari et al. 2016, Dehghan and Shah 2017). The goal here is to input possible additional knowledge associated with the scenes, detections, and other available information as the constraints to the model. Thus, the obtained tracks will be more accurate with respect to the ground truth. In this study, based on the assumption of using only one detector (body detector) and having one detection per person as inputs to the problem, we append the frame constraints in (37) to the basis model from Henschel et al. (2018), guaranteeing that no two detections inside a frame are associated with the same person.

There are a few works in the literature on object tracking which are related to our modeling and solution method. Leal-Taixe et al. (2012) study the problem of tracking multiple objects across multiple cameras. Their LP minimum cost flow formulation of the problem has block structural properties, and they explore the results using a branch-and-price algorithm. Wang et al. (2017) model the MOT problem through an ILP and suggest using CG for MOT and solving the associated pricing subproblem with dynamic programming. They consider a pool of constructed tracklets (short tracks

along with a few frames) as input for their model and choose an optimal among them based on DP. However, in this study, we deal with the quadratic interactions between many detections simultaneously.

4.3.1. Reformulation. To reformulate the problem as the general star-based model, (7)–(8), we identify each track j as the center of a possible star s. According to the definition of the problem, the pair-wise costs of MOT correspond only to the pairs of edges associated with the same person, meaning that if the edges are adjacent in H, their corresponding quadratic cost is nonzero and otherwise it is zero. Therefore, we can exploit the special structure of the adjacent-only class in this problem. Moreover, because the cost c_e is zero for the edges that are not incident to nodes in H, the objective function of our reformulation is reduced to the cost of stars. Thus, the star-based reformulation of MOT is given below:

$$[\text{RMP-MOT}]: \quad \min \quad \sum_{s\in\bar{S}} C_s\lambda_s \tag{39}$$

$$\text{s.t.} \quad \sum_{s\in\bar{S}} \lambda_s \le h \tag{40}$$

$$\sum_{s\in\bar{S}} D_{is}\lambda_s \le 1 \qquad \forall i \in N \tag{41}$$

$$\boldsymbol{\lambda} \in [0,1]^{|\bar{S}|} \tag{42}$$

The star-based model for LP relaxation of MOT consists of one star-only constraint (40) enforcing the maximum number of tracks, and one set of coupling constraints (41) to impose labeling every detection with at most one track.

4.3.2. Column Generation. The CG process starts from an empty set of feasible stars where we solve a pricing subproblem for each person $j \in H$ as a star center. Let π and $\rho_i, i \in N$, be the optimal solutions of dual variables associated with constraints (40) and (41), respectively, and by transferring the definition of ρ_i to the edge $e = \{i,j\}$, we rewrite it as ρ_e. We further define binary variable $z_e = 1$ when edge e is selected in the star and $z_e = 0$ otherwise. Given this, the pricing subproblem corresponding to center $j \in H$ is as follows:

$$\min \quad \sum_{e\in\delta(j)} (c_e - \rho_e)z_e + \sum_{e,f\in\delta(j)} q_{ef} z_e z_f - \pi \tag{43}$$

$$\text{s.t.} \quad \sum_{e\in\delta^t(j)} z_e \le 1 \qquad \forall t \in T \tag{44}$$

$$z_e \in \{0,1\} \qquad \forall e \in \delta(j) \tag{45}$$

where constraint (44) restricts each star to select a maximum of one detection per frame.

Note that this CG process requires only one subproblem to be solved at each iteration. This is because neither linear nor quadratic costs are dependent on the star centers; the centers can be realized identically. More specifically, suppose that $j, l \in H$ are possible star centers (tracks) and $i, k \in V \setminus H$ are the detections. The quadratic interaction between $e_1 = \{i,j\}$ and $f_1 = \{k,j\}$, $q_{e_1f_1}$ is equal to the quadratic interaction between $e_2 = \{i,l\}$ and $f_2 = \{k,l\}$, $q_{e_2f_2}$.

We remark that the pricing subproblem in this case is a constrained BQP problem. The implementation details are provided in Section 5.

5. Computational Experiments

In this section, we provide a rigorous experimental study to evaluate our suggested framework on test instances of the quadratic semi-assignment, the adjacent-only variant of the QSAP, and MOT in terms of dual bound and computing time. The generated data sets and results related to all experiments are available on GitHub.[1] We attempt to answer these fundamental questions through our experiments: How effective and applicable is the star-reformulation on other adjacent-only problems like the AQSAP and MOT? To what extent is employing the cost-splitting framework worthwhile for a BQP problem like the QSAP? And for each of the selected problems, which type of formulation performs better?

To answer these questions, we compare the star-reformulation, the cost-splitting, and the CG framework with both SLT and RLT as the two most commonly used methods in the literature of BQP. For the QSAP and the AQSAP, we also compare our results with the most effective exact method proposed (Rostami et al. 2023). GUROBI version 9.0.1 is chosen as our benchmark MIP solver, and we also solve the BQP models directly using GUROBI. Note that, through the CG procedure, we consider various possible reformulations which result in different pricing subproblems (i.e., BQP and different linearizations). This leads to different reformulations and solution strategies to solve the original

BQP models which can be implemented and solved by a commercial solver like GUROBI. We consider a time limit of three hours to solve each instance.

For the sake of comparison, the root node dual bound of the RMP is selected as a reliable indicator to be compared with the solver dual bound. Yet, the dual bound is not the only measure of efficiency; we single out the computation time needed to complete the CG process in the root node and for GUROBI to solve the problem.

We compare the trade-off between computation time and the quality of the dual bound using the methodology of performance profiles (Dolan and Moré 2002, Bergner et al. 2015) in the related section of each problem. The profile plot shows the cumulative distribution function (CDF) of the ratio of the time taken by each method to solve each problem or the bound obtained by each method to the best acquired time or bound among all the approaches for that problem. These two criteria are described below in more detail:

- **Dual bound performance profile:** The first set of graphs is based on the dual bound quality regardless of the time required to compute that bound. For each instance and each method, we compute the ratio between the dual bound of each method and the best bound among them. The horizontal axis reports this ratio; thus, the vertical axis corresponds to the fraction of instances with at least this ratio of bound performance displayed for each method. A large value is considered for the ratio where the method could not provide any dual bound for an instance within the time limit.

- **Time performance profile:** The second type of graph is generated based on the time needed to obtain the best dual bound. In this analysis, for each instance, we first find the best dual bound among all the dual bounds obtained by different methods. Then, for every method which yielded the best dual bound, we consider the ratio between the time required by that method to attain the best dual bound and the shortest time among all of them. This ratio provides the performance index in the horizontal axis. We report the fraction of instances with a maximum of a specific time ratio in the vertical axis. A large value is assigned to the ratio where the method is not able to achieve the best dual bound for an instance. Before moving on, let us make the following remarks:

Remark 1. For each method, the probability that the method will win over the rest of the methods is defined by the fraction of instances with the best performance (Dolan and Moré 2002). Hence, we refer to the ratio of instances with the performance equal to one as the number of wins of the associated method.

Remark 2. When running the CG algorithm, we do not have to wait until the termination of the CG procedure to obtain a valid lower bound. If the CG procedure could not converge in our desired time, we have information about the intermediate quality of the dual bound in each iteration at the expense of slightly more computations (Lübbecke and Desrosiers 2005). Nonetheless, to attain an exact dual bound, we must solve the pricing of that iteration to optimality.

Remark 3. Primal bounding methodologies are beyond the scope of the current study; thus, we do not embed the CG procedure in a branch-and-bound tree. However, we compute primal bounds by applying a trivial heuristic to the solutions of the CG for each instance of the problem. In this heuristic, we solve the IP model for the master problem of the last iteration where all available columns are considered as binary variables. In Online Appendix C, we discuss the obtained upper bounds in more detail.

Remark 4. The reported computation time for the experiments is comprised of the time to attain both LB and UB.

Remark 5. It is well accepted that difficulties in proving optimality may appear when column generation is solving a degenerate, large-scale problem. In addition, dual variables may oscillate from a good one to a much worse one, deriving the same value for many iterations of the CG (Amor et al. 2004, Desaulniers et al. 2006). Computational experiments show that it is possible to alleviate these effects using stabilization techniques such as dual-optimal inequalities and stabilized column generation algorithms. In this study, we implemented BoxPen, in-out separation, and interior-point stabilization techniques (Du Merle et al. 1999, Ben-Ameur and Neto 2007, Rousseau et al. 2007). Because these methods need tuning of several parameters and our main focus is on the application of the standard CG algorithm, further specialized CG enhancements and tuning procedures for specific problem structures were not thoroughly investigated. Our computational results demonstrate improvements in some CG experiments, but not for all of them. Thus, we have decided to keep the unstabilized results in our reports. However, one can test different techniques to accelerate the CG algorithm to solve the proposed star-reformulation and cost-splitting in an arbitrary application.

All the algorithms and models were implemented in the Python programming language. They were performed on a shared cluster with a four-core, 3.05-GHz processor and 128 GB RAM running under Linux 7.8. In the following sections, we present the test instances, parameter settings, and computational results for the data

association in MOT and both versions of the QSAP. The instance-by-instance tables in Online Appendix B provide more details on the results of each problem.

5.1. QSAP and Adjacent-Only QSAP Experiments

Instances in the literature that can be used for the QSAP experiments are very limited, and many papers rely on data instances that were randomly generated (Silva et al. 2021, Rostami et al. 2023). More importantly, because our main objective was to investigate the instances with varying cost structures, we generate random instances for QSAP, introduced in Section 4.1, where the processing cost c_{ij} for assigning a client i to a machine j is computed as $dm_i \times pr_j$. The client unit of demand for processing i is identified by dm_i, whereas pr_j indicates the required time for processing a unit in machine j. They are randomly generated over (0, 100) and (0, 10), respectively, from uniform distribution. We carry out our experiments considering different combinations of parameters $n \leq 50$ and $h \leq 14$, because in practice h is smaller than n. To investigate the impacts of the sparsity and structure of the quadratic matrix on the performance of the cost-splitting framework, we run the experiments on five randomly generated data sets with different sparsity of the quadratic matrices, each comprised of 41 instances. We embark on our experiments using an adjacent-only QSAP data set and adding out-of-star interactions incrementally to observe the effects. Indeed, the quadratic matrix of the first problem consists of in-star interactions only, whereas the next problems include 10%, 15%, 20%, and 25% out-of-star interactions, respectively, in addition to the in-star interactions. We should take into account that, in the real-life applications of BQP, the quadratic matrix is mostly very sparse (Furini et al. 2019). Therefore, adding just 10% out-of-star interactions on top of the in-star interactions results in a fairly dense quadratic matrix. Data generated by altering parameters n and h are defined in the QSAP-associated tables in Online Appendix B.

As mentioned in Section 4.1, the pricing subproblems (25) and (33) are unconstrained BQP problems. In order to heuristically solve these subproblems, we employ an open-source solver, qbsolv (Booth et al. 2017). Based on divide and conquer and dynamic programming, the solver partitions the problem into multiple subproblems and solves them using a tabu search algorithm. When the heuristic solver fails to find an improving column to add, we switch to an iteration of an exact method. In addition, as mentioned before, to retrieve information on the intermediate dual bounds, either the pricing has to be solved to optimality or a valid lower bound on the optimal reduced cost must be available. Hence, we apply a hybrid strategy to solve the pricing subproblem in which, after calling qbsolv for a fixed number of CG iterations, it switches to an exact method (branch-and-bound) for one iteration. In the exact iteration, either the BQP formulation of the subproblem or the linearized version is solved by GUROBI. We chose the standard linearization technique to construct the set $\mathcal{P}(x, y)$ in constraints (21). The concept of the standard linearization is presented in Online Appendix A.

As mentioned earlier, the RLT is, in general, the most effective linearization approach to provide tight bounds for the QSAP in the literature (Billionnet and Elloumi 2001, Schüle et al. 2009). Therefore, we compare the results of our reformulation with the results of GUROBI applied to the RLT model of the QSAP as well as the SLT model of the QSAP. To investigate the effectiveness of the proposed reformulation, we also implement the reformulation and outer approximation suggested in Rostami et al. (2023) for a class of BQP problems including the quadratic semi-assignment and report the results on the same instances. A brief description of the different reformulations and methods used to solve instances of the QSAP and the AQSAP is provided here.

BQP: The BQP model (14)–(16) solved by GUROBI.

SLT: Linearized reformulation (using SLT) of the BQP model (14)–(16) solved by GUROBI.

RLT: Linearized reformulation (using RLT) of the BQP model (14)–(16) solved by GUROBI.

OuterApproximation: Convex reformulation and outer approximation (Rostami et al. 2023) applied to the BQP model (14)–(16) solved by GUROBI.

CG+BQPPricing: CG algorithm for the model (17)–(23) where the UBQP pricing is solved by GUROBI.

CG+SLTPricing: CG algorithm for model (17)–(23) where the standard linearization of the pricing subproblem is solved by GUROBI.

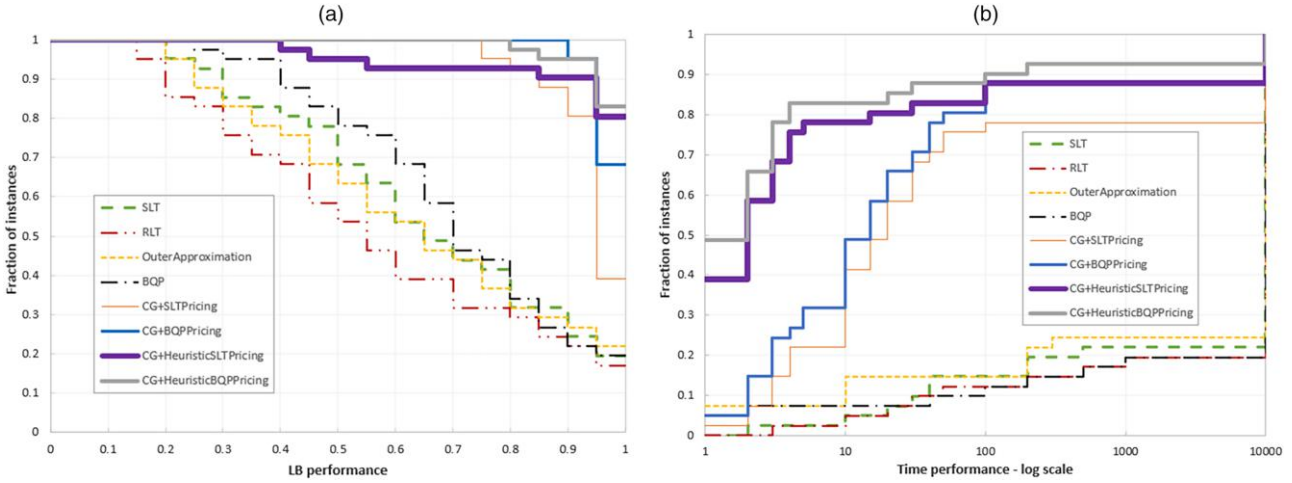
CG+HeuristicBQPPricing: CG algorithm for model (17)–(23) where the UBQP pricing is solved by using the hybrid heuristic method described above.

CG+HeuristicSLTPricing: CG algorithm for model (17)–(23) where the UBQP pricing is solved using the hybrid heuristic in which the exact iteration solves the standard linearized pricing subproblem by GUROBI.

Note that, although we use GUROBI to solve all the models obtained in different methods, in the rest of the work, we refer only to SLT, BQP, and RLT as GUROBI methods for the sake of simplicity.

Figures 3–7 show the results of all methods in terms of performance profile for both the AQSAP and the QSAP. In each figure, the left diagram compares all GUROBI, CG, and OuterApproximation methods in terms of dual bound performance, whereas the right one gives the time performance comparison of the methods. According to the description provided for the dual bound performance profile, the dual bound ratio is between zero and one for this

Figure 3. (Color online) Performance Profiles for AQSAP Instances

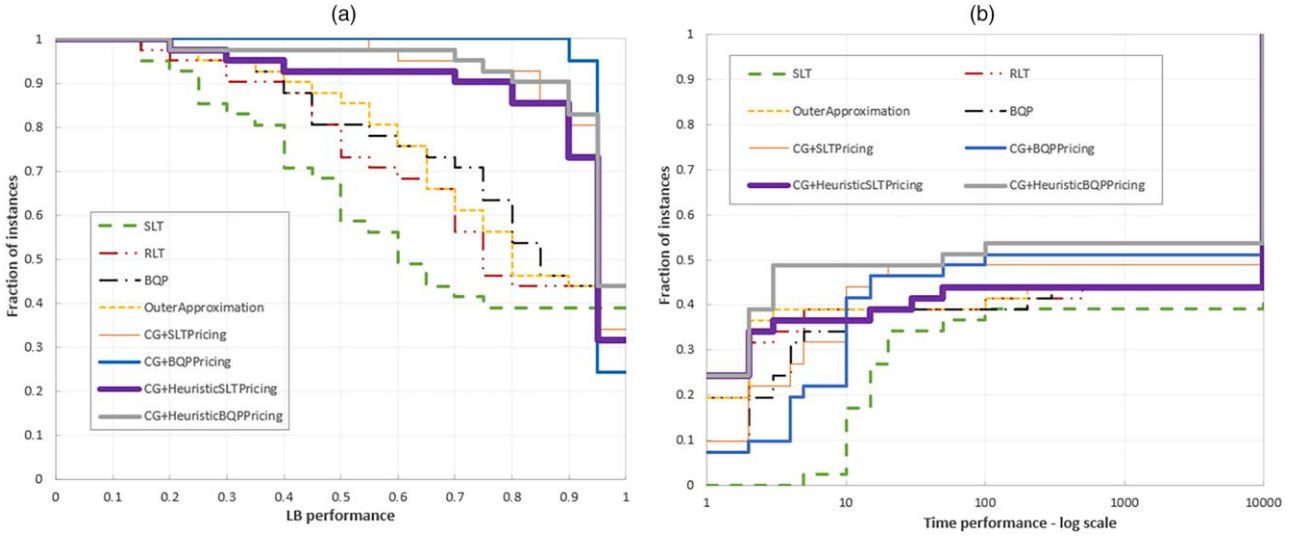


Notes. (a) Dual bound performance profile. (b) Time performance profile.

application. However, the diagrams on the right-hand side corresponding to time performance always consist of performance ratios that are greater than or equal to one. Clearly, for both bound and time performance profiles, a method with a larger fraction of instances with a ratio closer to one is preferable.

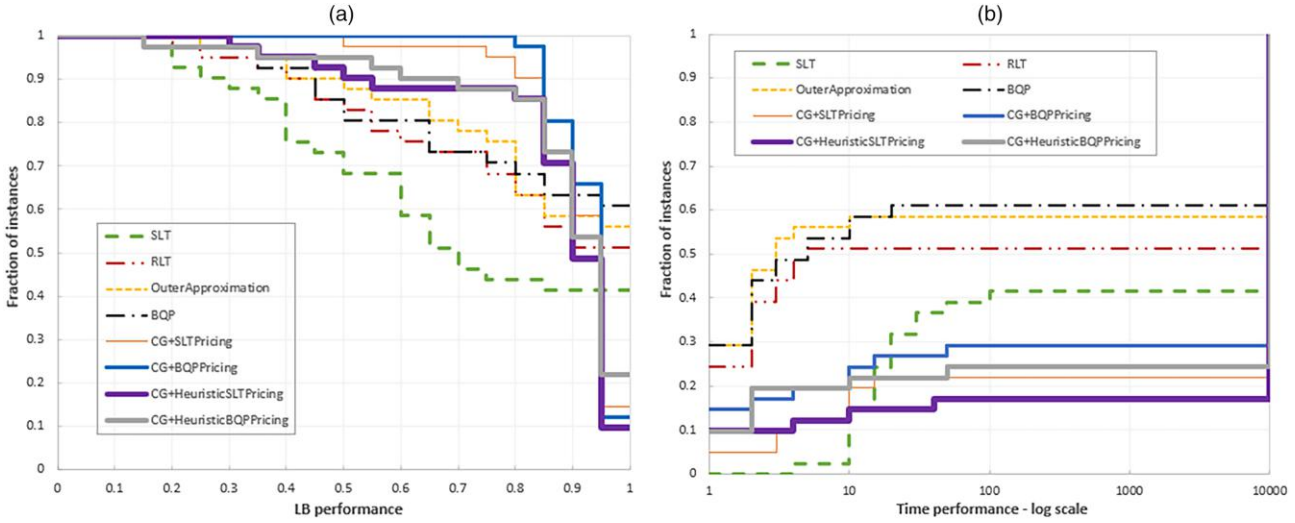
5.1.1. AQSAP Results. The results for the AQSAP instances are given in Figure 3. According to the LB performance analysis in Figure 3(a), CG hybrid methods have the most wins in terms of providing the best LB (83% and 80% for the CG+HeuristicBQPPricing and the CG+HeuristicSLTPricing, respectively) among all the methods, though the CG+BQPPricing outperforms the hybrid methods in a few quantiles. For instance, all the instances are solved by the CG+BQPPricing to 90% of the best LB, whereas only 95% of them could reach this ratio when solved by the CG+HeuristicPricing. We also observe that in general, GUROBI solves the BQP model (BQP) slightly better than it solves the linearization models (SLT and RLT). If we seek a method that can achieve at least 20% of the best LB, then all of the tested methods achieve this. However, if we increase the requirement to 40%, we can observe that all the CG methods perform better than the non-CG GUROBI methods and the OuterApproximation method. Finally, in looking for a method with 100% performance, the CG+HeuristicBQPPricing is the best choice.

Figure 4. (Color online) QSAP—10% Out-of-Star Density



Notes. (a) Dual bound performance profile. (b) Time performance profile.

Figure 5. (Color online) QSAP—15% Out-of-Star Density



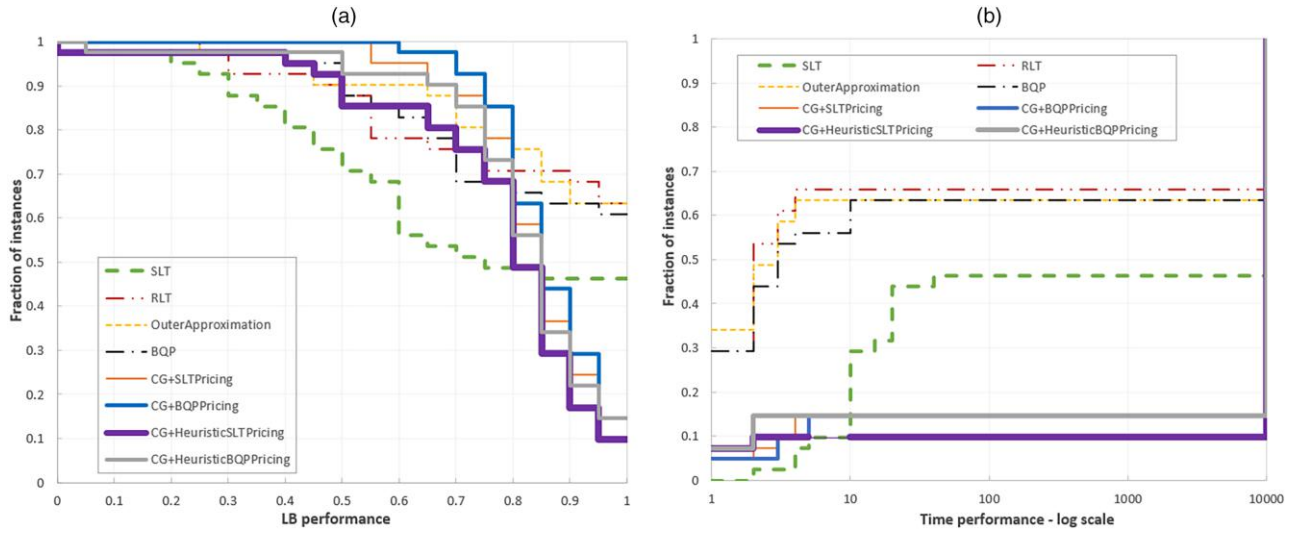
Notes. (a) Dual bound performance profile. (b) Time performance profile.

The next analyses are related to computing time. The time performance profile in Figure 3(b) shows that the best time to find the best LB for almost 50% of instances belongs to the CG+HeuristicBQPPricing method. It also demonstrates the superiority of this method in all other quantiles. Comparing all the methods together, the GUROBI methods and the OuterApproximation method are inefficient at finding the best LBs, where the best of them in terms of time (the OuterApproximation method) provides the best time in only 7% of the cases. As another example to illustrate the superiority of CG methods, consider that, using the SLT method, only 15% of instances and, using the BQP, only 12% of instances could converge to the best LB within two orders of magnitude of the best time.

Looking at both graphs simultaneously confirms the superiority of the CG-based methods over the GUROBI and the OuterApproximation methods, with a large gap for almost all intervals. Likewise, comparing the different methods of CG indicates that, as theory suggests, combining heuristics and exact methods to solve the pricing problems improves the results in terms of both computing time and LB for most of the instances.

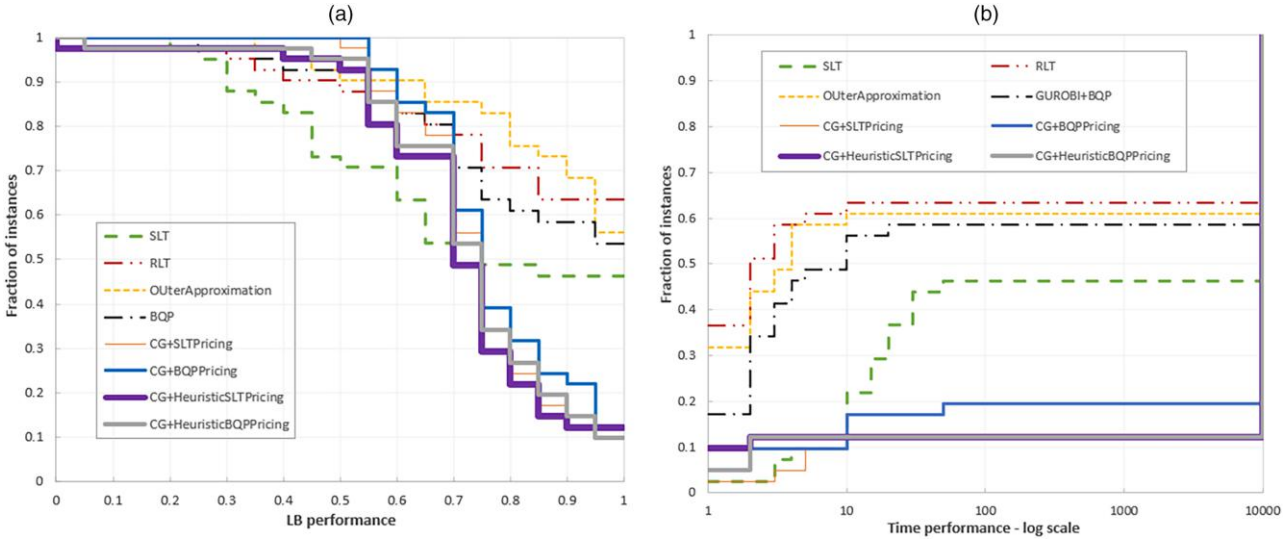
5.1.2. QSAP Results. The results for the QSAP instances are shown in Figures 4–7, where to increase the density of the quadratic matrix, we add more out-of-star interactions to the AQSAP each time to generate our data sets.

Figure 6. (Color online) QSAP—20% Out-of-Star Density



Notes. (a) Dual bound performance profile. (b) Time performance profile.

Figure 7. (Color online) QSAP—25% Out-of-Star Density



Notes. (a) Dual bound performance profile. (b) Time performance profile.

Considering the graph in Figure 4(a), where we have only 10% out-of-star interactions (quadratic costs), the overall interpretation of the results is very similar to adjacent-only QSAP. The CG+HeuristicBQPPricing has the greatest number of instances with the best LB among all the methods (44%); however, this time the CG+BQPPricing outperforms the hybrid counterpart in more intervals compared with the AQSAP. The performance of the GUROBI methods and the OuterApproximation is almost the same as before. The BQP and the OuterApproximation methods are superior to the other GUROBI methods, although they still have a huge LB performance gap compared with all the CG methods.

Nevertheless, in the next LB performance profile represented in Figure 5(a), all GUROBI methods and the OuterApproximation outperform CGs in the interval $[0.95, 1]$. Comparing the CG methods with one another in this figure demonstrates that the CG methods with exact pricing generally outperform heuristic pricing methods. Considering this figure and the next LB performance figures, Figures 6(a) and 7(a), the number of wins for the GUROBI and the OuterApproximation methods is more than for CG methods. Indeed, adding more out-of-star interactions incrementally to the data sets in Figures 6(a) and 7(a) results in enhancing the performance of the GUROBIs and the OuterApproximation method. However, even for the densest quadratic matrix in Figure 7(a), the CG methods still outperform non-CG methods in some quantiles. For example, we can observe the superiority of both CG+BQPPricing and CG+SLTPricing to GUROBI methods when a ratio smaller than 70% is considered and to the OuterApproximation method when the ratio is smaller than 60%. An overview of LB performance profiles for all of the data sets in the QSAP demonstrates that CG methods are more robust because they can achieve satisfactory LB performance for most of the instances, whereas if we need higher LB ratios, we aim to use GUROBI and the OuterApproximation. For instance, in the case of 20% out-of-star density and for a minimum requirement of 60% LB performance, we would opt for the CG+BQPPricing as all the instances reach this ratio of the best bound when they are solved using this method. In contrast, the best non-CG method, the OuterApproximation, satisfies this requirement in only 90% of the problem instances. Assessing non-CG methods, the plots suggest increasing the performance of both RLT and the OuterApproximation compared with BQP and SLT by adding more out-of-star costs, to an extent that they outperform BQP and SLT in most of the quantiles in Figure 7(a).

The second analysis is related to the time performance profiles. As shown in Figure 3(b), all CG methods perform better than GUROBI methods in all quantiles when it comes to AQSAP. Starting from the plot associated with time performance in Figure 4, when we add out-of-star interactions, the GUROBI methods and the OuterApproximation move upside of the figure and get closer to the CG methods. This is because the speed of the GUROBI and the OuterApproximation methods increases when the instances consist of some out-of-star interactions in addition to the in-star interactions. However, still, in Figure 4(b), most wins belong to CG methods. Starting from Figure 5(b), SLT, RLT, and BQP methods outperform all CG methods in all quantiles. Comparing non-CG methods together, similar to the LB analysis, the OuterApproximation outperforms the GUROBI methods in terms of time and it is superior even when we add more out-of-star costs.

It is interesting to note that, based on the last vertical lines of Figure 4(b), around 47% of the instances could not obtain the best LB when they are solved by the best CG method (CG+HeuristicBQPPricing), whereas this number is 57% for the best GUROBI method and the OuterApproximation method. Although inefficiency at finding the best LB increases for CG methods by adding more out-of-star costs, GUROBI methods can reach the best lower bound for more instances, in denser matrices, because of their speed improvement. Regarding the vertical lines for the BQP method in Figures 5(b) and 6(b), our interpretation is that this method could not obtain the best LB in the time limit for more than 35% of instances, whereas it obtained the best LB in the best time for the other 65% of instances. Moreover, it is observed that as often seen in the literature, CG methods with exact pricing are slower than their hybrid counterparts for all instances of adjacent-only problems. Nonetheless, the advantages of using the hybrid methods are weakened for more dense quadratic matrices. As an example, Figure 7(b) shows that the CG+HeuristicBQPPricing method outperforms the CG+BQPPricing in only 4% of the cases in terms of time performance. Looking at the trend in all of the time performance profiles in Figures 4(b)–7(b), we observe that the GUROBI methods and the OuterApproximation method generally outperform CG methods when the quadratic cost matrix of the QSAP is more dense. For instance, in Figure 7(b) where we have 25% out-of-star quadratic costs, in approximately 63% of the instances, the best GUROBI method (RLT) is within one order of magnitude with respect to the best time. Nevertheless, when we solve the instances using the best CG method (CG+BQPPricing), only around 18% of the cases are within this order. However, we should note that, in our time performance analysis, we do not reflect the computing time of the instances which could not obtain the best LB. In an extreme example, suppose that CG stopped in a few seconds with a high ratio of the best LB whereas GUROBI found the best LB in three hours. In this situation, because CG did not yield the best LB, we assign a very large number (10,000) to its time performance. If the GUROBI method is the only one that obtains the best LB, we consider its corresponding time performance equal to one.

Considering both performance profile analyses and the instance-by-instance details of Online Appendix B, it can be deduced that the star-based reformulation and CG, even with a basic implementation, outperform GUROBI in terms of computing time and LB for all the instances of the AQSAP. Additionally, in a large number of the instances, given a heuristic solution for the optimal value, optimality can be proved in the root node for this problem. By evaluating the general QSAP, the performance of the cost-splitting framework is reduced by augmenting the out-of-star quadratic matrix, and it might not be highly effective when the matrix is fairly dense. It should be noted that, because the majority of real applications of the BQP consist of sparse quadratic matrices (Furini et al. 2019), 10%–25% out-of-star interactions in addition to in-star interactions provide the proper condition to investigate the effects of having dense matrices in our proposed reformulation.

One notable remark is that, although the performance of the CG methods appears to deteriorate when solving the instances with more out-of-star costs, their performance in solving larger instances is still generally good. The results shown in instance-by-instance tables in Online Appendix B suggest that for the majority of large instances, the CG methods generally provide the best bounds when the time limit is reached. In this case, the GUROBI and OuterApproximation methods' dual bounds are relatively weak. In practice, we may choose CG methods over non-CG methods for even denser quadratic matrices. For example, when we have 15% out-of-star density, all the instances solved by the CG+BQPPricing provide an overall LB with the ratio of at least 80% of the best LB, whereas this ratio is 30% when using the best GUROBI method, BQP. Although GUROBI and OuterApproximation methods can solve small instances quite efficiently and obtain the best bounds, they still have a huge optimality gap when solving large instances after three hours. On the other hand, although CG methods are not able to converge to the best LB in many cases, they obtain a rather strong LB compared with the other methods in the majority of the instances which were not solved to optimality within the time limit.

5.2. MOT Experiments

For the MOT data association problem, we perform our tests based on a well-known benchmark, the MOT challenge data sets (Milan et al. 2016). Our tests are performed on different instances from the same video sequence (MOT16-09) of this benchmark and produce instances by altering the input parameters in the sequence. These parameters include the number of frames in the video ($\mathcal{T}$), the maximum number of tracks (h), and the maximum number of considered adjacent frames (d). Trivially, the estimated upper bound h has to be increased by enlarging the number of investigated data frames to avoid missing a track of any person in a real case. However, one has to consider the effects of this parameter on the growth of the problem size in the preestimation. In addition, we set the quadratic cost of two nodes that are more than d frames apart to zero in our implementation. We alter this parameter to demonstrate its influence on enlarging the quadratic matrix and problem size and generating various instances. The majority of the data set configurations in this section are based on Henschel et al. (2018).

It is worth noting that to be able to compare two entire MOT algorithms, the detection of objects and the precision of estimating unary and pair-wise costs, which are normally obtained using deep learning techniques, are very crucial.

However, because in the current paper we aim to compare the proposed reformulation of data association and CG with the results of a MIP solver, computing real accurate costs is beyond the scope of our research. Therefore, although we explore the real detection of the MOT challenge data set, we estimate naive unary and pair-wise costs for these detections based on some basic factors such as the distance between the detections. This method is exploited in several papers in data association. For example, Yarkony et al. (2020) assume that the costs are provided by learning methods and are given to their algorithms.

We should remark that each frame in the MOT16-09 sequence includes 15–25 detections; therefore, the number of decision variables in the represented formulation of (35) is between $15 \times h \times \mathcal{T}$ and $25 \times h \times \mathcal{T}$. To better realize the large scale of the problem, assume we investigate a data instance related to 10 frames of a video sequence consisting of an average of 20 detections in each frame and we aim to track a maximum number of 35 people in the sequence. This instance includes 7,000 decision variables in the represented BQP formulation, which generates a graph of the problem with 235 nodes in total. We show 27 generated instances from the mentioned data set in Tables 11 and 12 of Online Appendix B.

Similar to the semi-assignment problem, here we evaluate the efficiency of the star-based reformulation of MOT compared with the solver. In Section 4.3, we discussed that the pricing subproblem of MOT is a constrained BQP problem. So, we can apply a naturally tighter linearization, RLT, to solve the subproblems in addition to the previously used standard linearization in the QSAP. In Online Appendix A, we clarify this linearization when it is applied to the MOT formulation. Here, the solution methods are briefly introduced:

BQP: The BQP model (35)–(38) solved by GUROBI.

SLT: Linearized reformulation (using SLT) of the BQP model (35)–(38) solved by GUROBI.

RLT: Linearized reformulation (using RLT) of the BQP model (35)–(38) solved by GUROBI.

CG+BQPPricing: CG algorithm for model (39)–(42) where the constrained BQP pricing is solved by GUROBI.

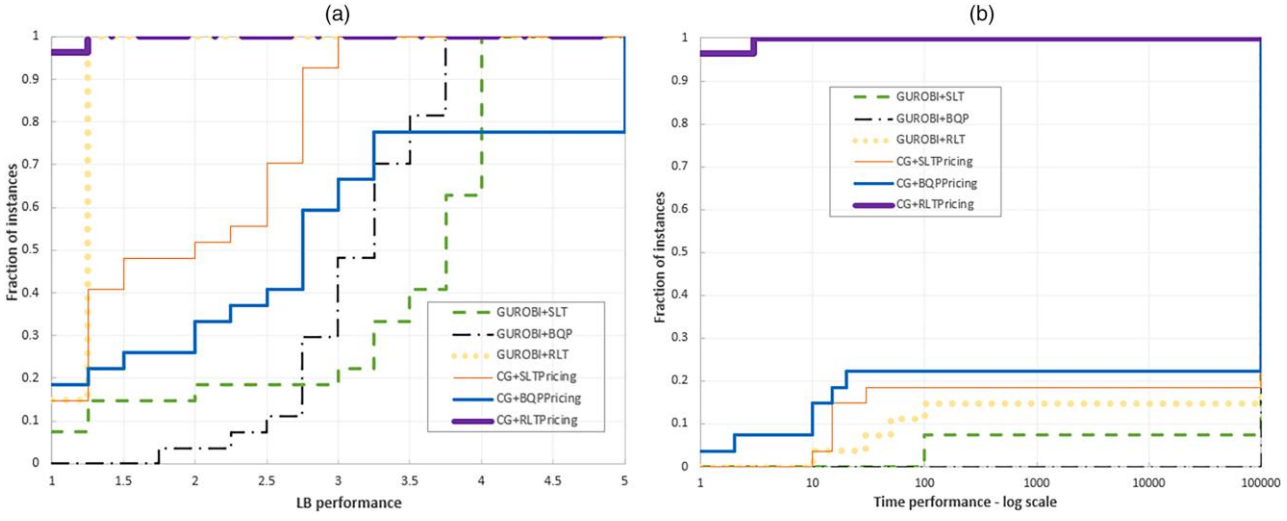
CG+SLTPricing: CG algorithm for the model (39)–(42) where the standard linearization of the pricing subproblem is solved by GUROBI.

CG+RLTPricing: CG algorithm for the model (39)–(42) where the linearized pricing subproblem using RLT is solved by GUROBI.

As discussed in the previous section, we investigate the performance of the methods through two types of performance profiles. Moreover, we report the experimental details for each instance of the problem in instance-by-instance tables in Online Appendix B.

According to the LB performance plot in Figure 8(a), all the CG-based methods outperform their GUROBI-based counterparts in almost all of the intervals. The only exception occurs when it comes to comparing the CG+BQPPricing method with the BQP method. The GUROBI method demonstrates superior performance in almost 20% of the instances. As we mentioned in the definition of the dual bound performance profile, we assign a large number for the performance ratio when the method fails to provide a valid dual bound for an instance within the time limit. Given that five is considered as a large ratio in our analyses, the figure indicates that in 22% of the instances, the

Figure 8. (Color online) Performance Profiles for the MOT Problem



Notes. (a) Dual bound performance profile. (b) Time performance profile.

CG+BQPPricing could not obtain an LB within the time limit. We identify the dual bound for these instances by "NA" in Table 12 in Online Appendix C. Another remark related to the LB performance graph is that all three CG methods have a number of wins that is equal to or greater than the best GUROBI method (RLT). Evidently, when we apply RLT to solve the pricing subproblem of the CG reformulation, the number of wins is the highest among all the methods with a large gap. Overall, the best method is the CG+RLTPricing, because it obtains the best LB for 96% of the instances and the worst LB outcome for this method for the rest of the instances is less than 1.25 times the best LB. It should be noted that the negative objective function of the data association problem in MOT is reflected in an LB performance ratio greater than one.

Although according to Figure 8(a) the ratio of the LB obtained by the RLT to the best LB is at most 1.25, Figure 8(b) shows that its computational time is not competitive compared with CGs. More specifically, less than 20% of the test set is solved by the RLT within two orders of magnitude with respect to the fastest method. Hence, it still outperforms other GUROBI methods. The performance profiles in Figure 8 delineate the superiority of the RLT method. Evidently, when the RLT is directly applied to the BQP formulation (the BQP+RLT method), it outperforms the other GUROBIs. Moreover, when it is used as the method for solving the pricing subproblem of CG (the CG+RLTPricing method), it has better performance than the other CG methods. Considering both LB and time performance, we observe that the CG+RLTPricing not only obtains the best LB in 96% of the cases but also does so in the shortest time for almost all of these cases. We detail the experiments on dual bound and time, as well as UB and the parameters of the instances, later in the Online Appendix.

Given the instance-by-instance tables in Online Appendix B and the performance analyses, we can infer that the star-based reformulation and the CG methodology computationally outperform the GUROBI solver in obtaining LB for the data association problem in MOT. Moreover, similar to the CG+RLTPricing, which outperforms the RLT, the other CG methods outperform their GUROBI counterparts in terms of both LB and computational time. Evidently, the CG+RLTPricing achieves the best LB in nearly all the cases, and, in the situation where it stops before the time limit, it converges to an optimal solution for the vast majority of instances.

6. Conclusion

In this study, we investigated the generalizability and performance of the star-reformulation on a large class of BQP problems, adjacent-only BQP problems. We employed the star-reformulation on two adjacent-only BQP problems with different characteristics: the quadratic semi-assignment problem and the multiple object tracking problem. Moreover, we developed a cost-splitting reformulation framework and solution methodology using column generation for general BQP problems. This framework, based on in-star and out-of-star interaction between pairs of edges in the BQP problem's graph, exploits the quadratic matrix structure. To evaluate the efficiency of our framework, we perform extensive experiments on the QSAP. We compare the lower bound and computing time of the reformulation and CG methods with a state-of-the-art MIP solver on instances of this problem as well as the AQSAP and MOT. According to the proposed framework, the adjacent-only class of problems inherits a special structure of the quadratic matrix which resulted in a huge improvement to solve these problems.

One notable outcome of this study is that, in the adjacent-only class of problems, a basic implementation of the presented framework can already compete with the solver, showing large improvements in terms of both dual bound and computation time. When out-of-star quadratic costs are added to the problem incrementally, the potential of the framework to compete with the solver decreases. Nevertheless, it is interesting to note that even in the case of QSAP with a fairly dense out-of-star quadratic matrix, the cost-splitting and CG methods still obtain promising results for some instances. Particularly in larger instances, where all the tested methods meet the time limit, CG methods outperform the MIP solver in many cases.

A possible future research direction is to explore the possibility of incorporating the proposed framework in a branch-and-bound tree to improve the primal bounds in addition to the dual bounds. Investigating this idea on other BQP problems with different constraint structures (such as the general case of quadratic minimum spanning tree) is another avenue for further research. Alternatively, exploring the star-reformulation idea on more problems in the adjacent-only BQP class such as the adjacent quadratic assignment problem is of interest for future studies. It would also be interesting to explore how effective the cost-splitting technique is on general BQP problems such as the quadratic assignment problem.

Acknowledgments

The authors thank the associate editor and anonymous reviewers for their valuable comments.

Endnote

[1] See https://github.com/mahbay/BqpDualBound.

References

Adams WP, Forrester RJ (2005) A simple recipe for concise mixed 0-1 linearizations. *Oper. Res. Lett.* 33(1):55–61.
Adams WP, Sherali HD (1990) Linearization strategies for a class of zero-one mixed integer programming problems. *Oper. Res.* 38(2):217–226.
Aloise D, Cafieri S, Caporossi G, Hansen P, Perron S, Liberti L (2010) Column generation algorithms for exact modularity maximization in networks. *Phys. Rev. E* 82(4):046112.
Amor HB, Desrosiers J, Frangioni A (2004) Stabilization in column generation. *Les Cahiers du GERAD* 711:2440.
Assad A, Xu W (1992) The quadratic minimum spanning tree problem. *Naval Res. Logist.* 39(3):399–417.
Assari SM, Idrees H, Shah M (2016) Human re-identification in crowd videos using personal, social and environmental constraints. Leibe B, Matas J, Sebe N, Welling M, eds. *Computer Vision–ECCV 2016.* Lecture Notes in Computer Science, vol. 9906 (Springer, Cham, Switzerland), 119–136.
Barahona F (1983) The max-cut problem on graphs not contractible to K5. *Oper. Res. Lett.* 2(3):107–111.
Ben-Ameur W, Neto J (2007) Acceleration of cutting-plane and column generation algorithms: Applications to network design. *Networks* 49(1):3–17.
Bergner M, Caprara A, Ceselli A, Furini F, Lübbecke ME, Malaguti E, Traversi E (2015) Automatic Dantzig-Wolfe reformulation of mixed integer programs. *Math. Programming* 149(1–2):391–424.
Bettiol E, Bomze I, Létocart L, Rinaldi F, Traversi E (2022) Mining for diamonds—Matrix generation algorithms for binary quadratically constrained quadratic problems. *Comput. Oper. Res.* 142(C):105735.
Billionnet A, Elloumi S (2001) Best reduction of the quadratic semi-assignment problem. *Discrete Appl. Math.* 109(3):197–213.
Billionnet A, Soutif É (2004) An exact method based on Lagrangian decomposition for the 0–1 quadratic knapsack problem. *Eur. J. Oper. Res.* 157(3):565–575.
Billionnet A, Elloumi S, Plateau MC (2009) Improving the performance of standard solvers for quadratic 0-1 programs by a tight convex reformulation: The QCR method. *Discrete Appl. Math.* 157(6):1185–1197.
Bonizzoni P, Della Vedova G, Dondi R, Jiang T (2008) On the approximation of correlation clustering and consensus clustering. *J. Comput. System Sci.* 74(5):671–696.
Booth M, Reinhardt SP, Roy A (2017) Partitioning optimization problems for hybrid classical/quantum execution. GitHub repository URL https://github.com/dwavesystems/qbsolv/blob/master/qbsolv_techReport.pdf.
Çela E (2013) *The Quadratic Assignment Problem: Theory and Algorithms*, vol. 1 (Springer Science & Business Media, Berlin).
Charfreitag J, Jünger M, Mallach S, Mutzel P (2022) McSparse: Exact solutions of sparse maximum cut and sparse unconstrained binary quadratic optimization problems. *2022 Proc. Sympos. Algorithm Engrg. Experiments (ALENEX)* (SIAM, Philadelphia), 54–66.
Chen WA, Zhu Z, Kong N (2018) A Lagrangian decomposition approach to computing feasible solutions for quadratic binary programs. *Optim. Lett.* 12:155–169.
Chrétienne P (1989) A polynomial algorithm to optimally schedule tasks on a virtual distributed system under tree-like precedence constraints. *Eur. J. Oper. Res.* 43(2):225–230.
Dantzig GB, Wolfe P (1960) Decomposition principle for linear programs. *Oper. Res.* 8(1):101–111.
De Fréminville PDLP, Desaulniers G, Rousseau LM, Perron S (2015) A column generation heuristic for districting the price of a financial product. *J. Oper. Res. Soc.* 66(6):965–978.
Dehghan A, Shah M (2017) Binary quadratic programing for online tracking of hundreds of people in extremely crowded scenes. *IEEE Trans. Pattern Anal. Machine Intelligence* 40(3):568–581.
Desaulniers G, Desrosiers J, Solomon MM (2006) *Column Generation*, vol. 5 (Springer Science & Business Media, Berlin).
Dolan ED, Moré JJ (2002) Benchmarking optimization software with performance profiles. *Math. Programming* 91(2):201–213.
Drwal M (2014) Algorithm for quadratic semi-assignment problem with partition size coefficients. *Optim. Lett.* 8(3):1183–1190.
Du Merle O, Villeneuve D, Desrosiers J, Hansen P (1999) Stabilized column generation. *Discrete Math.* 194(1):229–237.
Emami P, Pardalos PM, Elefteriadou L, Ranka S (2018) Machine learning methods for solving assignment problems in multi-target tracking. Preprint, submitted February 19, https://arxiv.org/abs/1802.06897.
Escoffier B, Hammer PL (2007) Approximation of the quadratic set covering problem. *Discrete Optim.* 4(3–4):378–386.
Fischer A (2014) An analysis of the asymmetric quadratic traveling salesman polytope. *SIAM J. Discrete Math.* 28(1):240–276.
Fischer F, Jaeger G, Lau A, Molitor P (2009) Complexity and algorithms for the traveling salesman problem and the assignment problem of second order. *Lecture Notes Comput. Sci.* 5165:211–224.
Furini F, Traversi E, Belotti P, Frangioni A, Gleixner A, Gould N, Liberti L, et al. (2019) QPLIB: A library of quadratic programming instances. *Math. Programming Comput.* 11(2):237–265.
Glover F, Woolsey E (1974) Converting the 0-1 polynomial programming problem to a 0-1 linear program. *Oper. Res.* 22(1):180–182.
Hahn PM, Zhu YR, Guignard M, Hightower WL, Saltzman MJ (2012) A level-3 reformulation-linearization technique-based bound for the quadratic assignment problem. *INFORMS J. Comput.* 24(2):202–209.
Hansen P, Lih KW (1992) Improved algorithms for partitioning problems in parallel, pipelined, and distributed computing. *IEEE Trans. Comput.* 41(6):769–771.
Helmberg C, Rendl F, Weismantel R (2000) A semidefinite programming approach to the Quadratic Knapsack Problem. *J. Combin. Optim.* 4(2):197–215.
Henschel R, Leal-Taixé L, Cremers D, Rosenhahn B (2018) Fusion of head and full-body detectors for multi-object tracking. *Proc. IEEE Conf. Comput. Vision Pattern Recognition Workshops* (IEEE, Piscataway, NJ), 1428–1437.
Hu H, Sotirov R (2018) Special cases of the quadratic shortest path problem. *J. Combin. Optim.* 35(3):754–777.
Hu H, Sotirov R (2021) The linearization problem of a binary quadratic problem and its applications. *Ann. Oper. Res.* 307(1):229–249.
Jünger M, Mallach S (2021) Exact facetial odd-cycle separation for maximum cut and binary quadratic optimization. *INFORMS J. Comput.* 33(4):1419–1430.
Khaniyev T (2018) Data-driven structure detection in optimization: Decomposition, hub location, and brain connectivity. PhD thesis, University of Waterloo, Waterloo, ON.
Khaniyev T, Elhedhli S, Erenay FS (2018) Structure detection in mixed-integer programs. *INFORMS J. Comput.* 30(3):570–587.

Khaniyev T, Elhedhli S, Erenay FS (2020) Spatial separability in hub location problems with an application to brain connectivity networks. *INFORMS J. Optim.* 2(4):320–346.
Kochenberger G, Hao JK, Glover F, Lewis M, Lü Z, Wang H, Wang Y (2014) The unconstrained binary quadratic programming problem: A survey. *J. Combin. Optim.* 28(1):58–81.
Leal-Taixe L, Pons-Moll G, Rosenhahn B (2012) Branch-and-price global optimization for multi-view multi-target tracking. *2012 IEEE Conf. Comput. Vision Pattern Recognition* (IEEE, Piscataway, NJ), 1987–1994.
Lemaréchal C, Oustry F (2001) SDP relaxations in combinatorial optimization from a Lagrangian viewpoint. Hadjisavvas N, Pardalos PM, eds. *Advances in Convex Analysis and Global Optimization.* Nonconvex Optimization and Its Applications, vol. 54 (Springer, Boston), 119–134.
Liberti L (2007) Compact linearization for binary quadratic problems. *4OR* 5(3):231–245.
Lübbecke ME, Desrosiers J (2005) Selected topics in column generation. *Oper. Res.* 53(6):1007–1023.
Magirou V, Milis J (1989) An algorithm for the multiprocessor assignment problem. *Oper. Res. Lett.* 8(6):351–356.
Mallach S (2018) Compact linearization for binary quadratic problems subject to assignment constraints. *4OR* 16(3):295–309.
Mallach S (2023) Inductive linearization for binary quadratic programs with linear constraints: A computational study. *4OR*, 1–41.
Malucelli F (1996) A polynomially solvable class of quadratic semi-assignment problems. *Eur. J. Oper. Res.* 91(3):619–622.
Mauri GR, Lorena LAN (2011) Lagrangean decompositions for the unconstrained binary quadratic programming problem. *Internat. Trans. Oper. Res.* 18(2):257–270.
Mauri GR, Lorena LAN (2012) A column generation approach for the unconstrained binary quadratic programming problem. *Eur. J. Oper. Res.* 217(1):69–74.
Meier JF, Clausen U, Rostami B, Buchheim C (2016) A compact linearisation of Euclidean single allocation hub location problems. *Electronic Notes Discrete Math.* 52:37–44.
Milan A, Leal-Taixe L, Reid I, Roth S, Schindler K (2016) MOT16: A benchmark for multi-object tracking. Preprint, submitted March 2, https://arxiv.org/abs/1603.00831.
O'Kelly ME (1987) A quadratic integer program for the location of interacting hub facilities. *Eur. J. Oper. Res.* 32(3):393–404.
Pereira DL, da Cunha AS (2018) Polyhedral results, branch-and-cut and Lagrangian relaxation algorithms for the adjacent only quadratic minimum spanning tree problem. *Networks* 71(1):31–50.
Pereira DL, da Cunha AS (2020) Dynamic intersection of multiple implicit Dantzig–Wolfe decompositions applied to the adjacent only quadratic minimum spanning tree problem. *Eur. J. Oper. Res.* 284(2):413–426.
Pereira DL, Gendreau M, Salles da Cunha A (2013) Stronger lower bounds for the quadratic minimum spanning tree problem with adjacency costs. *Electronic Notes Discrete Math.* 41(5):229–236.
Pereira DL, Gendreau M, Salles da Cunha A (2015) Branch-and-cut and branch-and-cut-and-price algorithms for the adjacent only quadratic minimum spanning tree problem. *Networks* 65(4):367–379.
Pisinger D (2007) The quadratic knapsack problem—A survey. *Discrete Appl. Math.* 155(5):623–648.
Punnen AP, Pandey P, Friesen M (2019) Representations of quadratic combinatorial optimization problems: A case study using quadratic set covering and quadratic knapsack problems. *Comput. Oper. Res.* 112:104769.
Punnen AP, Walter M, Woods BD (2017) A characterization of linearizable instances of the quadratic traveling salesman problem. Preprint, submitted August 23, https://arxiv.org/abs/1708.07217.
Rostami B, Malucelli F (2015) Lower bounds for the quadratic minimum spanning tree problem based on reduced cost computation. *Comput. Oper. Res.* 64:178–188.
Rostami B, Errico F, Lodi A (2023) A convex reformulation and an outer approximation for a large class of binary quadratic programs. *Oper. Res.* 71(2):471–486.
Rostami B, Malucelli F, Belotti P, Gualandi S (2016) Lower bounding procedure for the asymmetric quadratic traveling salesman problem. *Eur. J. Oper. Res.* 253(3):584–592.
Rostami B, Malucelli F, Frey D, Buchheim C (2015) On the quadratic shortest path problem. Bampis E, ed. *Proc. 14th Internat. Sympo. Experiment. Algorithms, SEA 2015* (Springer International Publishing, Cham, Switzerland), 379–390.
Rostami B, Chassein A, Hopf M, Frey D, Buchheim C, Malucelli F, Goerigk M (2018) The quadratic shortest path problem: Complexity, approximability, and solution methods. *Eur. J. Oper. Res.* 268(2):473–485.
Rousseau LM, Gendreau M, Feillet D (2007) Interior point stabilization for column generation. *Oper. Res. Lett.* 35(5):660–668.
Sahni S, Gonzalez T (1976) P-complete approximation problems. *J. ACM* 23(3):555–565.
Saito H, Fujie T, Matsui T, Matuura S (2009) A study of the quadratic semi-assignment polytope. *Discrete Optim.* 6(1):37–50.
Schüle I, Ewe H, Küfer KH (2009) Finding tight RLT formulations for quadratic semi-assignment problems. *Proc. 8th Cologne-Twente Workshop Graphs Combin. Optim.* (Ecole Polytechnique and CNAM, Paris), 109–112.
Shen H, Huang L, Huang C, Xu W (2018) Tracklet association tracker: An end-to-end learning-based association approach for multi-object tracking. Preprint, submitted August 5, https://arxiv.org/abs/1808.01562.
Sherali HD, Adams WP (2013) *A Reformulation-Linearization Technique for Solving Discrete and Continuous Nonconvex Problems.* Nonconvex Optimization and Its Applications, vol. 31 (Springer Science & Business Media, Berlin).
Sherali HD, Smith JC (2007) An improved linearization strategy for zero-one quadratic programming problems. *Optim. Lett.* 1(1):33–47.
Silva A, Coelho LC, Darvish M (2021) Quadratic assignment problem variants: A survey and an effective parallel memetic iterated tabu search. *Eur. J. Oper. Res.* 292(3):1066–1084.
Stone HS (1977) Multiprocessor scheduling with the aid of network flow algorithms. *IEEE Trans. Software Engrg.* SE-3(1):85–93.
Tang S, Andriluka M, Andres B, Schiele B (2017) Multiple people tracking by lifted multicut and person re-identification. *Proc. 30th IEEE Conf. Comput. Vision Pattern Recognition, CVPR 2017* (IEEE, Piscataway, NJ), 3539–3548.
Wang S, Wolf S, Fowlkes CC, Yarkony J (2017) Tracking objects with higher order interactions via delayed column generation. *Proc. 20th Internat. Conf. Artificial Intelligence Statist. (AISTATS 2017)* (PMLR, New York), 1132–1140.
Yarkony J, Adulyasak Y, Singh M, Desaulniers G (2020) Data association via set packing for computer vision applications. *INFORMS J. Optim.* 2(3):167–191.