ARTICLE IN PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Discrete Optimization ▮▮▮ (▮▮▮▮) ▮▮▮–▮▮▮

DISCRETE OPTIMIZATION

www.elsevier.com/locate/disopt

Comparisons and enhancement strategies for linearizing mixed 0-1 quadratic programs

Warren P. Adams[a,*], Richard J. Forrester[b], Fred W. Glover[c]

[a] *Clemson University, Clemson, SC 29634, USA*
[b] *Dickinson College, Carlisle, PA 17013, USA*
[c] *University of Colorado, Boulder, CO 80304, USA*

Received 24 December 2002; accepted 1 March 2004

Abstract

We present a linearization strategy for mixed 0-1 quadratic programs that produces small formulations with tight relaxations. It combines constructs from a classical method of Glover and a more recent reformulation-linearization technique (RLT). By using binary identities to rewrite the objective, a variant of the first method results in a concise formulation with the level-1 RLT strength. This variant is achieved as a modified surrogate dual of a Lagrangian subproblem to the RLT. Special structures can be exploited to obtain reductions in problem size, without forfeiting strength. Preliminary computational experience demonstrates the potential of the new representations.

Keywords: 0-1 Quadratic program; Reformulation; Linearization

1. Introduction

A standard practice in optimizing a mixed 0-1 quadratic program is to employ an initial linearization step that transforms the nonlinear problem into an equivalent linear form. For our purposes, two problems are said to be equivalent if they permit the same set of solutions in the original variable space and the objective function values equal at the corresponding solutions. The problem then becomes to optimize the resulting mixed 0-1 linear program. The motivation is to be able to solve the continuous relaxation of the linear form as a linear program so that a computationally inexpensive bound on the optimal objective function value to the nonlinear problem is available.

In order to achieve linearity, auxiliary variables and constraints are employed, with the newly defined variables replacing predesignated nonlinear expressions, and with the additional constraints enforcing that the new variables equal their nonlinear counterparts at all binary realizations of the 0-1 variables. The continuous relaxations of these representations tend to be repeatedly solved within enumerative frameworks as a means of fathoming nonoptimal or infeasible solutions. Of marked importance is that, although two different mixed 0-1 linear formulations may equivalently depict the same nonlinear problem, their sizes and continuous relaxations can drastically differ depending on the manner in which the auxiliary variables and constraints are defined. This leads to two key considerations of reformulation size and strength.

* Corresponding author.
E-mail addresses: wadams@clemson.edu (W.P. Adams), forrestr@dickinson.edu (R.J. Forrester), fred.glover@colorado.edu (F.W. Glover).

1572-5286/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.disopt.2004.03.006

From a computational point of view, there are tradeoffs between these considerations. The smaller formulations tend to promote inexpensive, though relatively weaker bounds. Certain larger representations are known to provide tighter bounds, though more effort is required to compute them. Generally speaking, formulations whose continuous relaxations provide tight approximations of the convex hull of solutions to the original nonlinear problem outperform the weaker representations. The "trick" is to obtain representations that balance the tradeoffs between size and strength so that effective bounds can be cheaply computed.

A classical linearization strategy that promotes very concise mixed 0-1 linear representations of mixed 0-1 quadratic programs is due to Glover [11]. Given such a problem having n binary variables, this method achieves linearity through the introduction of n unrestricted continuous variables and $4n$ linear inequalities. As shown in [1], a straightforward variant requires only n new nonnegative continuous variables and n new constraints. The problem conciseness results from the way in which each new continuous variable replaces the product of a binary variable and a linear function.

A more recent reformulation–linearization technique (RLT) of Sherali and Adams [23,24] is dedicated to obtaining formulations that promote tight approximations of discrete programs, with limited regard to problem size. The RLT provides for mixed 0-1 linear programs in n binary variables, an $(n+1)$-level hierarchy of progressively tighter polyhedral outer-approximations of the convex hull of solutions. These relaxations span the spectrum from the usual continuous relaxation at level 0 to the convex hull at level n. The RLT is identically applicable to quadratic programs, again providing a hierarchy of formulations. We focus in this paper on the level-1 formulation, which was originally applied to mixed 0-1 quadratic programs in the earlier works of [3,4], with computational experience reported in [5]. The strength of the RLT is due to the strategic manner in which the products of variables and constraints are computed, and in the substitution of a continuous variable for each product term.

A linearization of Lovász and Schrijver [20], when applied to pure 0-1 quadratic programs, produces the same representation as the level-1 RLT. Thus, certain relationships we will establish between [3,4] and [11] encompass [20] as well.

Returning to the method of Glover [11], depending on the manner in which the objective function to the original quadratic program is expressed, the strength of the continuous relaxation can vary. We show by first rewriting the objective function using simple binary identities, and then applying the idea of Glover to replace select nonlinear expressions with continuous variables, that concise formulations having the relaxation value of the level-1 RLT can be obtained. Thus we effectively combine the advantages of conciseness and strength within a single program.

Our analysis expresses a variant of [11] as a type of surrogate dual on a Lagrangian subproblem of the level-1 RLT representation; we first solve the level-1 RLT formulation as a linear program, and then use a subset of the optimal dual values to place specially designed equality restrictions into the objective function in such a manner that the subproblem has a block diagonal structure. These dualized constraints are the binary identities that define the rewritten objective function. The constraints in each subproblem block are then surrogated to obtain a variant of [11] with the strength of the level-1 RLT program. Two surrogate constraints per block ensure an equivalent linear representation. We further show how special structures in the constraints can be exploited to obtain reductions in problem size. These structures include set partitioning, variable upper bounding, and generalized upper bounding. Our computational experience indicates the overall promise of such an approach and, in particular, the utility of computing surrogates of the RLT constraints.

2. Mathematical background

We provide in this section limited mathematical background and notation that is needed to explain the research. In particular, we describe the linearization of [11] and the RLT of [23,24].

To establish notation, we present the general form of a mixed 0-1 quadratic program, referred to as Problem QP, below.

$$\begin{aligned} \text{QP : minimize} \quad & l(\boldsymbol{x},\boldsymbol{y}) + \sum_{j=1}^{n} g_j(\boldsymbol{x},\boldsymbol{y})x_j \\ \text{subject to} \quad & (\boldsymbol{x},\boldsymbol{y}) \in \boldsymbol{X} \equiv \{(\boldsymbol{x},\boldsymbol{y}) \in \boldsymbol{S} : \boldsymbol{x} \text{ binary}\} \end{aligned}$$

Here, $\boldsymbol{S}$ denotes a polyhedral set in the n discrete variables $\boldsymbol{x}$ and m continuous variables $\boldsymbol{y}$, and $l(\boldsymbol{x},\boldsymbol{y})$ and $g_j(\boldsymbol{x},\boldsymbol{y})$ for all j are linear functions in these same variables. We assume without loss of generality for each j that $g_j(\boldsymbol{x},\boldsymbol{y})$ is not a function of the variable x_j since $x_j^2 = x_j$ and that it does not contain a term of degree 0. Throughout, all indices run from 1 to n unless otherwise stated, the set $\boldsymbol{X}^R$ is used to denote any relaxation of $\boldsymbol{X}$ in the variables $(\boldsymbol{x},\boldsymbol{y})$, and the set $\boldsymbol{S}$ implies $\boldsymbol{0} \leqslant \boldsymbol{x} \leqslant \boldsymbol{1}$.

The methods of [11] and [23,24] are examined relative to Problem QP in Sections 2.1 and 2.2 respectively.

2.1. Glover's method

The procedure in [11] derives an equivalent mixed 0-1 linear representation of Problem QP by defining a new continuous variable z_j for each of the n products $g_j(\boldsymbol{x},\boldsymbol{y})x_j$ found in the objective function. It further introduces, for each j, four new inequalities to enforce that z_j equals $g_j(\boldsymbol{x},\boldsymbol{y})x_j$ at all binary realizations of $\boldsymbol{x}$. When applied to QP, Problem G results.

$$\begin{aligned} \text{G : minimize} \quad & l(\boldsymbol{x},\boldsymbol{y}) + \sum_{j=1}^{n} z_j \\ \text{subject to} \quad & L_j x_j \leqslant z_j \leqslant U_j x_j \quad \forall j && (1) \\ & g_j(\boldsymbol{x},\boldsymbol{y}) - U_j(1-x_j) \leqslant z_j \leqslant g_j(\boldsymbol{x},\boldsymbol{y}) - L_j(1-x_j) \quad \forall j && (2) \\ & (\boldsymbol{x},\boldsymbol{y}) \in \boldsymbol{X} \end{aligned}$$

As in [11], for each j, L_j and U_j are lower and upper bounds, respectively, on the linear functions $g_j(\boldsymbol{x},\boldsymbol{y})$ over $(\boldsymbol{x},\boldsymbol{y}) \in \boldsymbol{X}$. Such bounds can be calculated as

$$\begin{aligned} & L_j = \min\{g_j(\boldsymbol{x},\boldsymbol{y}) : (\boldsymbol{x},\boldsymbol{y}) \in \boldsymbol{X}^R\} \quad \text{and} \\ & U_j = \max\{g_j(\boldsymbol{x},\boldsymbol{y}) : (\boldsymbol{x},\boldsymbol{y}) \in \boldsymbol{X}^R\} && (3) \end{aligned}$$

where these problems are assumed bounded.

Inequalities (1) and (2) enforce the following equivalence between Problems QP and G: a point $(\boldsymbol{x},\boldsymbol{y})$ is feasible to Problem QP if and only if the point $(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})$ with $z_j = g_j(\boldsymbol{x},\boldsymbol{y})x_j$ for all j is feasible to Problem G with the same objective value. Given any $(\boldsymbol{x},\boldsymbol{y}) \in \boldsymbol{X}$, if some $x_j = 0$, then (1) ensures $z_j = 0$ with (2) redundant. If some $x_j = 1$, then (2) ensures $z_j = g_j(\boldsymbol{x},\boldsymbol{y})$ with (1) redundant. In either case, $z_j = g_j(\boldsymbol{x},\boldsymbol{y})x_j$ for each j.

Two simple observations lead to straightforward modifications of Problem G that reduce the problem size. First, since the intent is to use Problem G to compute an optimal solution to QP, the equivalence between these two problems need only hold at optimality. Consequently, we can eliminate the righthand inequalities of (1) and (2), and yet preserve the following equivalence: a point $(\boldsymbol{x},\boldsymbol{y})$ is *optimal* to Problem QP if and only if the point $(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})$ with $z_j = g_j(\boldsymbol{x},\boldsymbol{y})x_j$ for all j is *optimal* to Problem G with the same objective value. This observation was pointed out in [1], where it was also noted, provided $\boldsymbol{S} \subseteq \boldsymbol{X}^R$, that the optimal objective function value to the continuous relaxation of Problem G (obtained by removing the $\boldsymbol{x}$ binary restrictions) is unaffected by this removal of constraints. Second, using Glover [12], the number of structural constraints can be further reduced via either the substitution of variables $s_j = z_j - L_j x_j$ or $s_j = z_j - g_j(\boldsymbol{x},\boldsymbol{y}) + U_j(1-x_j)$ for each j. Such a substitution will replace n structural inequalities with the same number of nonnegativity restrictions, so that the overall procedure requires only n new nonnegative variables and n new structural constraints.

Before proceeding to Section 2.2 and reviewing the RLT procedure, we present below two enhancements to [11] that can tighten the continuous relaxation. The first demonstrates how to strengthen the bounds L_j and U_j computed in (3) and used in (1) and (2). The second introduces a rewrite of the objective function to QP using binary identities.

2.1.1. Enhancement 1: strengthening L_j and U_j

The bounds L_j and U_j computed in (3) can directly impact the optimal objective function value to the continuous relaxation of Problem G. We desire to increase the values of the lower bounds L_j and decrease the values of the upper bounds U_j to potentially tighten the continuous relaxation. To do so, we employ a conditional logic argument introduced in [26] and expanded in [19].

Let us begin with the lefthand inequalities of (1). For any given j, the associated inequality is essentially enforcing nonnegativity of the product of the nonnegative expressions x_j and $g_j(\boldsymbol{x},\boldsymbol{y}) - L_j$ as

$$x_j[g_j(\boldsymbol{x},\boldsymbol{y}) - L_j] \geqslant 0$$

where L_j is as defined in (3). The variable z_j in (1) replaces the quadratic term $x_j g_j(\boldsymbol{x},\boldsymbol{y})$ above. The concept of conditional logic applied to this quadratic inequality is that, since equality must hold under the condition that $x_j = 0$ regardless of the value of $g_j(\boldsymbol{x},\boldsymbol{y})$, we only need ensure that the second term in the expression is nonnegative when $x_j = 1$. Using this logic, we can replace the bound L_j with $L_j^1 = \min\{g_j(\boldsymbol{x},\boldsymbol{y}) : (\boldsymbol{x},\boldsymbol{y}) \in \boldsymbol{X}^R, x_j = 1\}$. An identical argument holds for the righthand inequalities of (1) since for each j the associated inequality can be viewed as

$$x_j[U_j - g_j(\boldsymbol{x},\boldsymbol{y})] \geqslant 0.$$

Here, the strengthened upper bound on $g_j(\boldsymbol{x},\boldsymbol{y})$, say U_j^1, can be computed as in (3) with the additional restriction that $x_j = 1$.

Similarly, by observing for each *j* that the righthand and lefthand inequalities in (2) can be obtained by enforcing nonnegativity of the products of the nonnegative expressions $1 - x_j$ with each of $g_j(\boldsymbol{x}, \boldsymbol{y}) - L_j$ and $U_j - g_j(\boldsymbol{x}, \boldsymbol{y})$ respectively, we obtain that the corresponding bounds L_j and U_j can be analogously tightened, this time under the conditional logic restriction that $x_j = 0$. We use the notation L_j^0 and U_j^0 to represent these new bounds. The net result is to reformulate Problem G as G2 below.

$$
\begin{aligned}
\text{G2 : minimize} \quad & l(\boldsymbol{x}, \boldsymbol{y}) + \sum_{j=1}^{n} z_j \\
\text{subject to} \quad & L_j^1 x_j \leqslant z_j \leqslant U_j^1 x_j \quad \forall j && (4) \\
& g_j(\boldsymbol{x}, \boldsymbol{y}) - U_j^0(1 - x_j) \leqslant z_j \leqslant g_j(\boldsymbol{x}, \boldsymbol{y}) - L_j^0(1 - x_j) \quad \forall j && (5) \\
& (\boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{X}
\end{aligned}
$$

Here,

$$
\begin{aligned}
& L_j^1 = \min\{g_j(\boldsymbol{x}, \boldsymbol{y}) : (\boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{X}^R, x_j = 1\} \quad \text{and} \\
& U_j^1 = \max\{g_j(\boldsymbol{x}, \boldsymbol{y}) : (\boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{X}^R, x_j = 1\} && (6)
\end{aligned}
$$

and

$$
\begin{aligned}
& L_j^0 = \min\{g_j(\boldsymbol{x}, \boldsymbol{y}) : (\boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{X}^R, x_j = 0\} \quad \text{and} \\
& U_j^0 = \max\{g_j(\boldsymbol{x}, \boldsymbol{y}) : (\boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{X}^R, x_j = 0\}. && (7)
\end{aligned}
$$

By definition we have that $L_j^1 \geqslant L_j$, $L_j^0 \geqslant L_j$, $U_j^1 \leqslant U_j$, and $U_j^0 \leqslant U_j$ for each *j*. Of course, if for some *j* the problems in either (6) or (7) have no solution then the variable x_j can be fixed to a binary value in QP, with QP infeasible if both (6) and (7) have no solution.

Since Problem G2 affords a potentially tighter relaxation than G without additional effort, the remainder of this paper will focus on comparisons to G2. We note that although the righthand inequalities of (1) and (2) of Problem G can be eliminated without altering the optimal objective function value to G or its continuous relaxation, provided $\boldsymbol{S} \subseteq \boldsymbol{X}^R$, the analogous argument for G2 does not hold. While binary equivalence between Problems QP and G2 will continue to hold when the righthand inequalities of (4) and (5) of G2 are eliminated, the continuous relaxation of G2 could be weakened.

The example below demonstrates that the relaxation of Problem G2 can give a tighter bound than that of G, and that the removal of the righthand inequalities of (4) and (5) can weaken the continuous relaxation of G2 (though never beyond the relaxation value of G).

Example 2.1. Consider the following instance of Problem QP having $n = 2$ binary variables $\boldsymbol{x}$ and no continuous variables $\boldsymbol{y}$ so that the functions $l(\boldsymbol{x}, \boldsymbol{y})$, $g_1(\boldsymbol{x}, \boldsymbol{y})$, and $g_2(\boldsymbol{x}, \boldsymbol{y})$ reduce to $l(\boldsymbol{x})$, $g_1(\boldsymbol{x})$, and $g_2(\boldsymbol{x})$, respectively.

$$
\begin{aligned}
\text{minimize} \quad & 3x_1 - 3x_2 + (-1x_2)x_1 + (0x_1)x_2 \\
\text{subject to} \quad & \boldsymbol{x} \in \boldsymbol{X} \equiv \{\boldsymbol{x} \in \boldsymbol{S} = \{(x_1, x_2) : 2x_1 - 2x_2 \geqslant -1, -x_1 + x_2 \geqslant 0, x_1 \geqslant 0, \\
& \qquad -x_2 \geqslant -1\} : x_1, x_2 \text{ binary}\}
\end{aligned}
$$

Thus, $l(\boldsymbol{x}) = 3x_1 - 3x_2$, $g_1(\boldsymbol{x}) = -1x_2$, and $g_2(\boldsymbol{x}) = 0x_1$ in QP. We first compute the bounds $(L_1, L_2) = (-1, 0)$ and $(U_1, U_2) = (0, 0)$ as prescribed in (3) with $\boldsymbol{X}^R = \boldsymbol{S}$, and then construct Problem G. The optimal objective value to the continuous relaxation of G is -2. Next form Problem G2 by computing the bounds $(L_1^1, L_2^1) = (-1, 0)$ and $(U_1^1, U_2^1) = (-1, 0)$ as in (6), and $(L_1^0, L_2^0) = (-\frac{1}{2}, 0)$ and $(U_1^0, U_2^0) = (0, 0)$ as in (7), again using $\boldsymbol{X}^R = \boldsymbol{S}$. The optimal objective value to the continuous relaxation of G2 is -1.5, which exceeds the value -2 obtained using G. However, if we eliminate the righthand inequalities of (4) and (5) in G2, the optimal objective value to the continuous relaxation of G2 is weakened to -2.

2.1.2. Enhancement 2: rewriting the objective function

The manner in which the objective function to Problem QP is expressed can affect the relaxation value of Problem G2. Indeed, even a minor adjustment such as the recording of a quadratic term $x_i x_j$ as $x_j x_i$ can alter the value. The below example demonstrates such an alteration.

Example 2.2. Consider the following instance of Problem QP having $n = 4$ binary variables $\boldsymbol{x}$ and no continuous variables $\boldsymbol{y}$.

$$\begin{array}{ll} \text{minimize} & -4x_1 + x_2 + x_4 + (5x_2 - x_3 - 2x_4)x_1 + (-2x_3)x_2 + (x_4)x_3 + (0)x_4 \\ \text{subject to} & \boldsymbol{x} \in \boldsymbol{X} \equiv \{\boldsymbol{x} \in \boldsymbol{S} = \{\boldsymbol{x} : \boldsymbol{0} \leqslant \boldsymbol{x} \leqslant \boldsymbol{1}\} : \boldsymbol{x} \text{ binary}\} \end{array}$$

The functions $l(\boldsymbol{x})$ and $g_j(\boldsymbol{x})$ for $j = 1, \ldots, 4$ are accordingly: $l(\boldsymbol{x}) = -4x_1 + x_2 + x_4$, $g_1(\boldsymbol{x}) = 5x_2 - x_3 - 2x_4$, $g_2(\boldsymbol{x}) = -2x_3$, $g_3(\boldsymbol{x}) = x_4$, and $g_4(\boldsymbol{x}) = 0$. Programs (6) and (7) give $(L_1^1, L_2^1, L_3^1, L_4^1) = (L_1^0, L_2^0, L_3^0, L_4^0) = (-3, -2, 0, 0)$ and $(U_1^1, U_2^1, U_3^1, U_4^1) = (U_1^0, U_2^0, U_3^0, U_4^0) = (5, 0, 1, 0)$ so that Problem G2 (and also Problem G for this instance) becomes the below.

$$\begin{array}{ll} \text{minimize} & -4x_1 + x_2 + x_4 + z_1 + z_2 + z_3 + z_4 \\ \text{subject to} & -3x_1 \leqslant z_1 \leqslant 5x_1 \\ & -2x_2 \leqslant z_2 \leqslant 0x_2 \\ & 0x_3 \leqslant z_3 \leqslant 1x_3 \\ & 0x_4 \leqslant z_4 \leqslant 0x_4 \\ & 5x_2 - x_3 - 2x_4 - 5(1 - x_1) \leqslant z_1 \leqslant 5x_2 - x_3 - 2x_4 + 3(1 - x_1) \\ & -2x_3 - 0(1 - x_2) \leqslant z_2 \leqslant -2x_3 + 2(1 - x_2) \\ & x_4 - 1(1 - x_3) \leqslant z_3 \leqslant x_4 - 0(1 - x_3) \\ & 0 - 0(1 - x_4) \leqslant z_4 \leqslant 0 - 0(1 - x_4) \\ & \boldsymbol{0} \leqslant \boldsymbol{x} \leqslant \boldsymbol{1}, \boldsymbol{x} \text{ binary} \end{array}$$

(Observe that $z_4 = 0$ at all feasible solutions so that this variable could have been eliminated from the problem.) The optimal objective function value to the continuous relaxation is $-\frac{21}{4}$, with an optimal solution $(x_1, x_2, x_3, x_4, z_1, z_2, z_3, z_4) = (\frac{3}{4}, 0, 1, 0, -\frac{9}{4}, 0, 0, 0)$.

If we add the quantity $\frac{5}{2}(x_1x_2 - x_2x_1)$ to the objective function so that the coefficient on x_2 in $g_1(\boldsymbol{x})$ decreases to $\frac{5}{2}$ and the coefficient of x_1 in $g_2(\boldsymbol{x})$ increases to $\frac{5}{2}$, we get $U_1^1 = U_1^0 = U_2^1 = U_2^0 = \frac{5}{2}$, with all other lower and upper bounds unchanged. The continuous relaxation to the resulting linearization has the optimal objective function value -5 with optimal solutions $(x_1, x_2, x_3, x_4, z_1, z_2, z_3, z_4) = (1, 0, 1, 0, -1, 0, 0, 0)$, $(1, 0, 0, 1, -2, 0, 0, 0)$, and $(1, 0, 1, 1, -3, 0, 1, 0)$. As they are integral, these points are also optimal to Problem QP.

In light of the above example, the question arises as to how best express the objective function to QP before applying the method of [11]. In fact, we can also consider quadratic terms that involve complements $\bar{x}_j$ of the binary variables x_j, where $\bar{x}_j = 1 - x_j$. Specifically, suppose we add multiples of the binary identities

$$x_i x_j = x_j x_i \quad \forall (i, j), i < j \tag{8}$$

$$x_i \bar{x}_j = x_i - x_i x_j \quad \forall (i, j), i \neq j \tag{9}$$

$$y_i \bar{x}_j = y_i - y_i x_j \quad \forall (i, j), i = 1, \ldots, m \tag{10}$$

to the objective function using suitably dimensioned vectors $\boldsymbol{\alpha}^1$, $\boldsymbol{\alpha}^2$, and $\boldsymbol{\alpha}^3$, respectively, to obtain an equivalent problem to QP of the below form:

$$\begin{array}{ll} \text{QP}(\boldsymbol{\alpha}) : \text{minimize} & l^{\boldsymbol{\alpha}}(\boldsymbol{x}, \boldsymbol{y}) + \sum_{j=1}^{n} g_j^{\boldsymbol{\alpha}}(\boldsymbol{x}, \boldsymbol{y}) x_j + \sum_{j=1}^{n} h_j^{\boldsymbol{\alpha}}(\boldsymbol{x}, \boldsymbol{y}) \bar{x}_j \\ \qquad\text{subject to} & (\boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{X} \end{array}$$

where

$$l^{\boldsymbol{\alpha}}(\boldsymbol{x}, \boldsymbol{y}) = l(\boldsymbol{x}, \boldsymbol{y}) + \sum_{j=1}^{n} \left(\sum_{\substack{i=1 \\ i \neq j}}^{n} \alpha_{ij}^2 x_i + \sum_{i=1}^{m} \alpha_{ij}^3 y_i \right), \tag{11}$$

$$g_j^{\boldsymbol{\alpha}}(\boldsymbol{x}, \boldsymbol{y}) = g_j(\boldsymbol{x}, \boldsymbol{y}) - \sum_{i=1}^{j-1} \alpha_{ij}^1 x_i + \sum_{i=j+1}^{n} \alpha_{ji}^1 x_i - \sum_{\substack{i=1 \\ i \neq j}}^{n} \alpha_{ij}^2 x_i - \sum_{i=1}^{m} \alpha_{ij}^3 y_i \quad \forall j \tag{12}$$

and

$$h_j^{\boldsymbol{\alpha}}(\boldsymbol{x}, \boldsymbol{y}) = -\sum_{\substack{i=1 \\ i \neq j}}^{n} \alpha_{ij}^2 x_i - \sum_{i=1}^{m} \alpha_{ij}^3 y_i \quad \forall j. \tag{13}$$

The basic premise of [11] can be applied to each quadratic expression $g_j^{\boldsymbol{\alpha}}(\boldsymbol{x}, \boldsymbol{y})x_j$ and $h_j^{\boldsymbol{\alpha}}(\boldsymbol{x}, \boldsymbol{y})\bar{x}_j$. Of course, in order to linearize the newly introduced expressions $h_j^{\boldsymbol{\alpha}}(\boldsymbol{x}, \boldsymbol{y})\bar{x}_j$, an additional n continuous variables and $4n$ inequalities beyond the method of [11] are employed. As we will see in Section 3, however, the resulting formulations afford very tight linear programming bounds that relate to the level-1 RLT relaxation value, and certain of these $4n$ inequalities can be removed from consideration. Interestingly, Section 4.1 identifies special structures for which these additional variables and constraints are not needed to achieve the level-1 relaxation strength. For now, let us replace the quadratic expressions $g_j^{\boldsymbol{\alpha}}(\boldsymbol{x}, \boldsymbol{y})x_j$, and $h_j^{\boldsymbol{\alpha}}(\boldsymbol{x}, \boldsymbol{y})\bar{x}_j$ with continuous variables z_j^1 and z_j^2, respectively, and define $8n$ linear inequalities to ensure that each of these variables z_j^1 and z_j^2 equals their respective quadratic expression at all binary realizations of $\boldsymbol{x}$. The problem below emerges.

$$\begin{aligned}
\mathrm{G2}(\boldsymbol{\alpha}) : \text{minimize} \quad & l^{\boldsymbol{\alpha}}(\boldsymbol{x}, \boldsymbol{y}) + \sum_{j=1}^{n} z_j^1 + \sum_{j=1}^{n} z_j^2 \\
\text{subject to} \quad & L_j^{\boldsymbol{\alpha}1} x_j \leqslant z_j^1 \leqslant U_j^{\boldsymbol{\alpha}1} x_j \quad \forall j && (14) \\
& g_j^{\boldsymbol{\alpha}}(\boldsymbol{x}, \boldsymbol{y}) - U_j^{\boldsymbol{\alpha}0}(1 - x_j) \leqslant z_j^1 \leqslant g_j^{\boldsymbol{\alpha}}(\boldsymbol{x}, \boldsymbol{y}) - L_j^{\boldsymbol{\alpha}0}(1 - x_j) \quad \forall j && (15) \\
& \bar{L}_j^{\boldsymbol{\alpha}0}(1 - x_j) \leqslant z_j^2 \leqslant \bar{U}_j^{\boldsymbol{\alpha}0}(1 - x_j) \quad \forall j && (16) \\
& h_j^{\boldsymbol{\alpha}}(\boldsymbol{x}, \boldsymbol{y}) - \bar{U}_j^{\boldsymbol{\alpha}1} x_j \leqslant z_j^2 \leqslant h_j^{\boldsymbol{\alpha}}(\boldsymbol{x}, \boldsymbol{y}) - \bar{L}_j^{\boldsymbol{\alpha}1} x_j \quad \forall j && (17) \\
& (\boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{X}
\end{aligned}$$

Here, for each j, the values $L_j^{\boldsymbol{\alpha}1}$ and $U_j^{\boldsymbol{\alpha}1}$ are computed as in (6) as

$$\begin{aligned}
L_j^{\boldsymbol{\alpha}1} &= \min\{g_j^{\boldsymbol{\alpha}}(\boldsymbol{x}, \boldsymbol{y}) : (\boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{X}^R, x_j = 1\} \quad \text{and} \\
U_j^{\boldsymbol{\alpha}1} &= \max\{g_j^{\boldsymbol{\alpha}}(\boldsymbol{x}, \boldsymbol{y}) : (\boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{X}^R, x_j = 1\}
\end{aligned} \tag{18}$$

while the values $L_j^{\boldsymbol{\alpha}0}$ and $U_j^{\boldsymbol{\alpha}0}$ are computed as in (7) as

$$\begin{aligned}
L_j^{\boldsymbol{\alpha}0} &= \min\{g_j^{\boldsymbol{\alpha}}(\boldsymbol{x}, \boldsymbol{y}) : (\boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{X}^R, x_j = 0\} \quad \text{and} \\
U_j^{\boldsymbol{\alpha}0} &= \max\{g_j^{\boldsymbol{\alpha}}(\boldsymbol{x}, \boldsymbol{y}) : (\boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{X}^R, x_j = 0\}.
\end{aligned} \tag{19}$$

Similarly, for each j, the values $\bar{L}_j^{\boldsymbol{\alpha}1}$ and $\bar{U}_j^{\boldsymbol{\alpha}1}$ are computed as

$$\begin{aligned}
\bar{L}_j^{\boldsymbol{\alpha}1} &= \min\{h_j^{\boldsymbol{\alpha}}(\boldsymbol{x}, \boldsymbol{y}) : (\boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{X}^R, x_j = 1\} \quad \text{and} \\
\bar{U}_j^{\boldsymbol{\alpha}1} &= \max\{h_j^{\boldsymbol{\alpha}}(\boldsymbol{x}, \boldsymbol{y}) : (\boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{X}^R, x_j = 1\}
\end{aligned} \tag{20}$$

with the values $\bar{L}_j^{\boldsymbol{\alpha}0}$ and $\bar{U}_j^{\boldsymbol{\alpha}0}$ computed as

$$\begin{aligned}
\bar{L}_j^{\boldsymbol{\alpha}0} &= \min\{h_j^{\boldsymbol{\alpha}}(\boldsymbol{x}, \boldsymbol{y}) : (\boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{X}^R, x_j = 0\} \quad \text{and} \\
\bar{U}_j^{\boldsymbol{\alpha}0} &= \max\{h_j^{\boldsymbol{\alpha}}(\boldsymbol{x}, \boldsymbol{y}) : (\boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{X}^R, x_j = 0\}.
\end{aligned} \tag{21}$$

The notation $\mathrm{QP}(\boldsymbol{\alpha})$ and $\mathrm{G2}(\boldsymbol{\alpha})$, and the superscript $\boldsymbol{\alpha}$ used throughout these problems as well as in (18)–(21), are to denote their dependence on the values of $\boldsymbol{\alpha} = (\boldsymbol{\alpha}^1, \boldsymbol{\alpha}^2, \boldsymbol{\alpha}^3)$. We elected to substitute $\bar{x}_j = 1 - x_j$ for all j so that the variables $\bar{x}_j$ do not appear in (16), (17), (20), or (21).

Regardless of the chosen values of $\boldsymbol{\alpha}$, the mixed 0-1 linear program $\mathrm{G2}(\boldsymbol{\alpha})$ is equivalent to the quadratic program QP, with the optimal objective value to the continuous relaxation of $\mathrm{G2}(\boldsymbol{\alpha})$, say $v(\boldsymbol{\alpha})$, providing a lower bound on the optimal objective value to QP. The task is to determine an $\boldsymbol{\alpha}$ that provides the maximum possible lower bound. That is, we wish to solve the nonlinear program

$$\text{NP}: v^* = \max_{\boldsymbol{\alpha}} v(\boldsymbol{\alpha}) = \min \left\{ l^{\boldsymbol{\alpha}}(\boldsymbol{x}, \boldsymbol{y}) + \sum_{j=1}^{n} z_j^1 + \sum_{j=1}^{n} z_j^2 : (14)\text{–}(17), (\boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{S} \right\}. \tag{22}$$

In Section 3, we solve Problem NP by comparing it to the level-1 RLT formulation of [23,24], reviewed in the following section.

2.2. The RLT

The RLT produces, for mixed 0-1 linear and polynomial programs, a hierarchy of successively tighter linear programming approximations. At each level of the hierarchy, the linear problem is equivalent to the nonlinear program when the $\boldsymbol{x}$ binary restrictions are enforced, but yields a relaxation when the binary restrictions are weakened to $\mathbf{0} \leqslant \boldsymbol{x} \leqslant \mathbf{1}$. At the highest level n, where n represents the number of binary variables, the linear program is exact in that the feasible region gives an explicit description of the convex hull of solutions to the nonlinear program, with the linear objective function equalling the original nonlinear objective at each extreme point solution. Consequently, at this highest level, the $\boldsymbol{x}$ binary restrictions can be equivalently replaced by $\mathbf{0} \leqslant \boldsymbol{x} \leqslant \mathbf{1}$.

The RLT consists of the two basic steps of *reformulation* and *linearization*. The reformulation step generates redundant, nonlinear inequalities by multiplying the problem constraints by product factors of the binary variables and their complements, recognizing and enforcing that $x_j^2 = x_j$ for each binary variable x_j. The linearization step recasts the problem into a higher variable space by replacing each distinct product with a continuous variable. The hierarchical levels are defined in terms of the product factors employed, with the individual levels dependent on the degrees of these factors. We concern ourselves in this paper with the (weakest) level-1 formulations, originally appearing in [3,4]. For a thorough description of the basic RLT theory, the reader is referred to [23,24], with a detailed overview of the various applications and extensions in [25].

Let us construct the level-1 RLT representation of Problem QP. Suppose, without loss of generality, that the polyhedral set $\boldsymbol{S}$ is given by

$$\boldsymbol{S} = \left\{ (\boldsymbol{x}, \boldsymbol{y}) : \sum_{i=1}^{n} a_{ri} x_i + \sum_{i=1}^{m} d_{ri} y_i \geqslant b_r \;\; \forall r = 1, \ldots, R \right\} \tag{23}$$

and that the linear functions $g_j(\boldsymbol{x}, \boldsymbol{y})$ for all j are expressed as follows:

$$g_j(\boldsymbol{x}, \boldsymbol{y}) = \sum_{\substack{i=1 \\ i \neq j}}^{n} C_{ij} x_i + \sum_{i=1}^{m} D_{ij} y_i \quad \forall j = 1, \ldots, n. \tag{24}$$

The reformulation step multiplies each inequality defining $\boldsymbol{S}$ by each binary variable x_j and its complement $(1 - x_j)$ for all $j = 1, \ldots, n$, substituting throughout $x_j^2 = x_j$ for all j. The linearization step then substitutes a continuous variable for each product in the objective function and constraints, in this case letting $w_{ij}^1 = x_i x_j$ for all $i = 1, \ldots, n$, $i \neq j$, and $\gamma_{ij}^1 = y_i x_j$ for all $i = 1, \ldots, m$. We choose here to implement additional substitutions found within [23,24]. In particular, we let $w_{ij}^2 = x_i - w_{ij}^1$ for all $i = 1, \ldots, n$, $i \neq j$, and $\gamma_{ij}^2 = y_i - \gamma_{ij}^1$ for all $i = 1, \ldots, m$ throughout each constraint which was multiplied by a $(1 - x_j)$ factor, and then explicitly enforce these substitutions as constraints. Clearly, we have that $w_{ij}^1 = w_{ji}^1$ for all (i, j), $i < j$, and so these restrictions are also enforced, resulting in the following program:

$$\text{QPRLT}: \text{minimize} \quad l(\boldsymbol{x}, \boldsymbol{y}) + \sum_{j=1}^{n} \left(\sum_{\substack{i=1 \\ i \neq j}}^{n} C_{ij} w_{ij}^1 + \sum_{i=1}^{m} D_{ij} \gamma_{ij}^1 \right)$$

$$\text{subject to} \quad \sum_{\substack{i=1 \\ i \neq j}}^{n} a_{ri} w_{ij}^1 + \sum_{i=1}^{m} d_{ri} \gamma_{ij}^1 \geqslant (b_r - a_{rj}) x_j \quad \forall (r, j), \;\; r = 1, \ldots, R \tag{25}$$

$$\sum_{\substack{i=1 \\ i \neq j}}^{n} a_{ri} w_{ij}^2 + \sum_{i=1}^{m} d_{ri} \gamma_{ij}^2 \geqslant b_r (1 - x_j) \quad \forall (r, j), \;\; r = 1, \ldots, R \tag{26}$$

$$w^1_{ij} = w^1_{ji} \quad \forall (i, j),\ i < j \tag{27}$$

$$w^2_{ij} = x_i - w^1_{ij} \quad \forall (i, j),\ i \neq j \tag{28}$$

$$\gamma^2_{ij} = y_i - \gamma^1_{ij} \quad \forall (i, j),\ i = 1, \ldots, m \tag{29}$$

$$(\boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{S} \tag{30}$$

$$\boldsymbol{x} \text{ binary}$$

Inequalities (25) result from multiplying the constraints of $\boldsymbol{S}$ by x_j for each j while inequalities (26) result from multiplying these same constraints by each $(1 - x_j)$ and making the substitutions of (28) and (29).

The RLT theory enforces at all feasible solutions to QPRLT that $w^1_{ij} = x_i x_j$ and $w^2_{ij} = x_i(1 - x_j)$ for all $(i, j), i \neq j$, and that $\gamma^1_{ij} = y_i x_j$ and $\gamma^2_{ij} = y_i(1 - x_j)$ for all $(i, j), i = 1, \ldots, m$. As alluded to above, the level-1 RLT formulation [23,24] does not need to explicitly include constraints (27) through (29), nor the variables w^1_{ij} for all $(i, j), i > j$, w^2_{ij} for all $(i, j), i \neq j$, and γ^2_{ij} for all $(i, j), i = 1, \ldots, m$. Instead, the substitutions suggested by these constraints can be performed to eliminate the corresponding variables, making the restrictions themselves unnecessary. In addition, inequalities (30) are unnecessary as they are implied by (25), (26), (28), and (29). We choose here to consider the larger form given by QPRLT, as the additional variables and constraints facilitate our arguments in the upcoming section.

3. Combining conciseness and strength

The main result of this section is that the optimal objective function values to Problem NP and the continuous relaxation of Problem QPRLT equal, and that an optimal value of $\boldsymbol{\alpha}$ for NP can be obtained from any optimal dual solution to QPRLT, using the multipliers corresponding to constraints (27)–(29). This will hold true provided that the set $\boldsymbol{X}^R$ used to compute bounds (18)–(21), and found in (14)–(17), is defined as the set $\boldsymbol{S}$, which we henceforth assume. We also assume for each $j = 1, \ldots, n$ that $\min\{x_j : \boldsymbol{x} \in \boldsymbol{S}\} = 0$ and $\max\{x_j : \boldsymbol{x} \in \boldsymbol{S}\} = 1$ since otherwise variables can be accordingly fixed to binary values. The significance of this result is that the strength of the level-1 RLT formulation can be captured in a program having the concise size of G2($\boldsymbol{\alpha}$).

Certain notation is adopted for convenience. Consistent with the construction of Problem QPRLT, let the expressions $\lfloor g^{\boldsymbol{\alpha}}_j(\boldsymbol{x}, \boldsymbol{y}) x_j \rfloor_L$ and $\lfloor h^{\boldsymbol{\alpha}}_j(\boldsymbol{x}, \boldsymbol{y}) x_j \rfloor_L$ denote, for each j, the linearized forms of the products $g^{\boldsymbol{\alpha}}_j(\boldsymbol{x}, \boldsymbol{y}) x_j$ and $h^{\boldsymbol{\alpha}}_j(\boldsymbol{x}, \boldsymbol{y}) x_j$, respectively, obtained by substituting $w^1_{ij} = x_i x_j$ for all $i \neq j$, and $\gamma^1_{ij} = y_i x_j$ for all $i = 1, \ldots, m$ so that

$$\begin{aligned}\lfloor g^{\boldsymbol{\alpha}}_j(\boldsymbol{x}, \boldsymbol{y}) x_j \rfloor_L = {} & \sum_{\substack{i=1 \\ i \neq j}}^{n} C_{ij} w^1_{ij} + \sum_{i=1}^{m} D_{ij} \gamma^1_{ij} \\ & - \sum_{i=1}^{j-1} \alpha^1_{ij} w^1_{ij} + \sum_{i=j+1}^{n} \alpha^1_{ji} w^1_{ij} - \sum_{\substack{i=1 \\ i \neq j}}^{n} \alpha^2_{ij} w^1_{ij} - \sum_{i=1}^{m} \alpha^3_{ij} \gamma^1_{ij} \quad \forall j\end{aligned} \tag{31}$$

by (12) and (24), and

$$\lfloor h^{\boldsymbol{\alpha}}_j(\boldsymbol{x}, \boldsymbol{y}) x_j \rfloor_L = -\sum_{\substack{i=1 \\ i \neq j}}^{n} \alpha^2_{ij} w^1_{ij} - \sum_{i=1}^{m} \alpha^3_{ij} \gamma^1_{ij} \quad \forall j \tag{32}$$

by (13). Consequently, since the linearization operation gives

$$\lfloor g^{\boldsymbol{\alpha}}_j(\boldsymbol{x}, \boldsymbol{y}) x_j \rfloor_L = g^{\boldsymbol{\alpha}}_j(\boldsymbol{x}, \boldsymbol{y}) - \lfloor g^{\boldsymbol{\alpha}}_j(\boldsymbol{x}, \boldsymbol{y})(1 - x_j) \rfloor_L \quad \forall j, \tag{33}$$

we have by substituting (28) and (29) into (31) that

$$\lfloor g_j^{\boldsymbol{\alpha}}(\mathbf{x},\mathbf{y})(1-x_j)\rfloor_L = \sum_{\substack{i=1\\ i\neq j}}^{n} C_{ij}w_{ij}^2 + \sum_{i=1}^{m} D_{ij}\gamma_{ij}^2 - \sum_{i=1}^{j-1}\alpha_{ij}^1 w_{ij}^2 + \sum_{i=j+1}^{n}\alpha_{ji}^1 w_{ij}^2 - \sum_{\substack{i=1\\ i\neq j}}^{n}\alpha_{ij}^2 w_{ij}^2 - \sum_{i=1}^{m}\alpha_{ij}^3\gamma_{ij}^2 \quad \forall j. \tag{34}$$

Similarly, since

$$\lfloor h_j^{\boldsymbol{\alpha}}(\mathbf{x},\mathbf{y})x_j\rfloor_L = h_j^{\boldsymbol{\alpha}}(\mathbf{x},\mathbf{y}) - \lfloor h_j^{\boldsymbol{\alpha}}(\mathbf{x},\mathbf{y})(1-x_j)\rfloor_L \quad \forall j, \tag{35}$$

we have by substituting (28) and (29) into (32) that

$$\lfloor h_j^{\boldsymbol{\alpha}}(\mathbf{x},\mathbf{y})(1-x_j)\rfloor_L = -\sum_{\substack{i=1\\ i\neq j}}^{n}\alpha_{ij}^2 w_{ij}^2 - \sum_{i=1}^{m}\alpha_{ij}^3\gamma_{ij}^2 \quad \forall j. \tag{36}$$

For each j, the notation $\lfloor g_j^{\boldsymbol{\alpha}}(\hat{\mathbf{x}},\hat{\mathbf{y}})\hat{x}_j\rfloor_L$ and $\lfloor h_j^{\boldsymbol{\alpha}}(\hat{\mathbf{x}},\hat{\mathbf{y}})\hat{x}_j\rfloor_L$ is used to denote the values $\lfloor g_j^{\boldsymbol{\alpha}}(\mathbf{x},\mathbf{y})x_j\rfloor_L$ and $\lfloor h_j^{\boldsymbol{\alpha}}(\mathbf{x},\mathbf{y})x_j\rfloor_L$ at the point $(\hat{\mathbf{x}},\hat{\mathbf{y}},\hat{\mathbf{w}},\hat{\boldsymbol{\gamma}})$ as prescribed by (31) and (32), respectively.

We use this notation in the proof of the below theorem. This theorem formally states the dominance of the level-1 RLT representation relative to Problem G2($\boldsymbol{\alpha}$).

Theorem 1. *The optimal objective function value to the continuous relaxation of Problem QPRLT is an upper bound on the optimal objective value to the relaxation of* G2($\boldsymbol{\alpha}$), *regardless of the chosen* $\boldsymbol{\alpha}$.

Proof. Arbitrarily select a vector $\boldsymbol{\alpha}$. It is sufficient to show, using obvious vector notation, that given any feasible solution $(\hat{\mathbf{x}},\hat{\mathbf{y}},\hat{\mathbf{w}},\hat{\boldsymbol{\gamma}})$ to the continuous relaxation of QPRLT, the point $(\hat{\mathbf{x}},\hat{\mathbf{y}},\hat{\mathbf{z}})$ having $\hat{z}_j^1 = \lfloor g_j^{\boldsymbol{\alpha}}(\hat{\mathbf{x}},\hat{\mathbf{y}})\hat{x}_j\rfloor_L$ for all j and $\hat{z}_j^2 = \lfloor h_j^{\boldsymbol{\alpha}}(\hat{\mathbf{x}},\hat{\mathbf{y}})(1-\hat{x}_j)\rfloor_L$ for all j is feasible to the relaxation of G2($\boldsymbol{\alpha}$) with the same objective function value. Toward this end, for each j, twice surrogate inequalities (25), once each with an optimal set of dual multipliers to the minimization and maximization problems in (18), to verify by (31) that $(\hat{\mathbf{x}},\hat{\mathbf{y}},\hat{\mathbf{z}})$ satisfies the lefthand and righthand inequalities, respectively, of (14). Similarly, for each j, twice surrogate inequalities (26), once each with an optimal set of dual multipliers to the minimization and maximization problems in (21), to verify by (36) that $(\hat{\mathbf{x}},\hat{\mathbf{y}},\hat{\mathbf{z}})$ satisfies the lefthand and righthand inequalities, respectively, of (16). In an analogous manner, again twice surrogate the inequalities (25), once each with optimal dual multipliers to the optimization problems in (20) to verify by (32) and (35) that (17) is satisfied, and twice surrogate inequalities (26), once each using optimal dual multipliers to (21) to verify by (34) and (33) that (15) is satisfied. Hence $(\hat{\mathbf{x}},\hat{\mathbf{y}},\hat{\mathbf{z}})$ is feasible to G2($\boldsymbol{\alpha}$).The objective function value to G2($\boldsymbol{\alpha}$) at this point is $l^{\boldsymbol{\alpha}}(\hat{\mathbf{x}},\hat{\mathbf{y}}) + \sum_{j=1}^{n}(\lfloor g_j^{\boldsymbol{\alpha}}(\hat{\mathbf{x}},\hat{\mathbf{y}})\hat{x}_j\rfloor_L + \lfloor h_j^{\boldsymbol{\alpha}}(\hat{\mathbf{x}},\hat{\mathbf{y}})(1-\hat{x}_j)\rfloor_L)$, which equals the objective value to QPRLT at $(\hat{\mathbf{x}},\hat{\mathbf{y}},\hat{\mathbf{w}},\hat{\boldsymbol{\gamma}})$ since the former is by definition obtained by adding constraints (27), (28), and (29) to the objective function of QPRLT using multipliers $\boldsymbol{\alpha}^1$, $\boldsymbol{\alpha}^2$, and $\boldsymbol{\alpha}^3$, respectively. This completes the proof. □

In order to establish our desired result equating the optimal objective function values to Problems NP and the continuous relaxation of QPRLT, with an optimal $\boldsymbol{\alpha}$ to NP consisting of a partial optimal dual vector to QPRLT, we construct a Lagrangian dual to this latter problem. In particular, we place constraints (27)–(29) into the objective function using the multipliers $\boldsymbol{\alpha}=(\boldsymbol{\alpha}^1,\boldsymbol{\alpha}^2,\boldsymbol{\alpha}^3)$. Incorporating the notation of (11), (31), and (36), Problem LD results.

LD : maximize $\theta(\boldsymbol{\alpha})$

where

$$\theta(\boldsymbol{\alpha}) = \min\left\{ l^{\boldsymbol{\alpha}}(\mathbf{x},\mathbf{y}) + \sum_{j=1}^{n}\lfloor g_j^{\boldsymbol{\alpha}}(\mathbf{x},\mathbf{y})x_j\rfloor_L + \sum_{j=1}^{n}\lfloor h_j^{\boldsymbol{\alpha}}(\mathbf{x},\mathbf{y})(1-x_j)\rfloor_L : (25),\ (26),\ \text{and } (30)\right\} \tag{37}$$

Our argument is based on a special block-diagonal structure that the Lagrangian subproblem $\theta(\boldsymbol{\alpha})$ possesses. This structure was our reason for explicitly including constraints (27), (28), and (29) in QPRLT, as opposed to substituting out the variables w_{ij}^1 for all $(i,j), i>j$, w_{ij}^2 for all $(i,j), i\neq j$, and γ_{ij}^2 for all (i,j), and then removing these restrictions. Indeed, $\theta(\boldsymbol{\alpha})$ has $2n$

separate blocks: one block over each of constraints (25) and (26) for each j, coupled by the restrictions $(\boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{S}$ found in (30). The theorem below shows that this structure can be exploited to efficiently compute $\theta(\boldsymbol{\alpha})$ by solving a linear program whose objective function is expressed in terms of the parameters $L_j^{\boldsymbol{\alpha}1}$ and $\bar{L}_j^{\boldsymbol{\alpha}0}$ of (18) and (21), and whose constraints are the coupling restrictions $(\boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{S}$.

Theorem 2. *Given any vector $\boldsymbol{\alpha}$, the value $\theta(\boldsymbol{\alpha})$ in* (37) *is equal to the optimal objective function value of the linear program*

$$\text{LP}(\boldsymbol{\alpha}): \quad \text{minimize} \left\{ l^{\boldsymbol{\alpha}}(\boldsymbol{x}, \boldsymbol{y}) + \sum_{j=1}^{n} L_j^{\boldsymbol{\alpha}1} x_j + \sum_{j=1}^{n} \bar{L}_j^{\boldsymbol{\alpha}0}(1 - x_j) : (\boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{S} \right\}, \tag{38}$$

where for each j, $L_j^{\boldsymbol{\alpha}1}$ and $\bar{L}_j^{\boldsymbol{\alpha}0}$ are computed as in (18) *and* (21), *respectively.*

Proof. The proof is to show for each j that an optimal set of dual multipliers to the corresponding inequalities in (25) of LD can be computed using any optimal dual solution to the minimization problem in (18) and that an optimal set of dual multipliers to the corresponding inequalities in (26) of LD can be computed using any optimal dual solution to the minimization problem in (21). The result must then hold since the dual to Problem $\text{LP}(\boldsymbol{\alpha})$ in (38) is the dual to the minimization problem of (37), where the multipliers to constraints (25) and (26) of LD have been fixed in the former at an optimal set of values.

Suppose for a given j that we solve the minimization problem in (18) to obtain a primal optimal solution, and denote it by $\tilde{w}_{ij}^1$ for all $i \neq j$ and $\tilde{\gamma}_{ij}^1$ for all i to represent the x_i and y_i variables, respectively. Further suppose that we fix the dual multipliers to the associated constraints in (25) equal to the computed optimal duals to (18). Similarly, suppose we solve the minimization problem in (21) to obtain a primal optimal solution, and denote it by $\tilde{w}_{ij}^2$ for all $i \neq j$ and $\tilde{\gamma}_{ij}^2$ for all i to represent the x_i and y_i variables, respectively. Further suppose that we fix the dual multipliers to the associated constraints in (26) equal to the computed optimal duals to (21). Repeating for each j we obtain dual multipliers for all the constraints (25) and (26). Solve the dual to Problem LD with these fixed dual values, which necessarily satisfy dual feasibility relative to the w_{ij}^1, w_{ij}^2, γ_{ij}^1, and γ_{ij}^2 variables, to obtain an $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) \in \boldsymbol{S}$ and multipliers $\hat{\boldsymbol{\zeta}}$. The fixed duals for (25) and (26) together with $\hat{\boldsymbol{\zeta}}$ define a dual feasible solution to LD since dual feasibility relative to the variables x_i and y_i are ensured by solving the reduced dual to Problem LD. Moreover, for the same reason, $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ and $\hat{\boldsymbol{\zeta}}$ satisfy complementary slackness relative to (30) since they are optimal primal and dual solutions, respectively, to this same problem. Finally, $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{w}}, \hat{\boldsymbol{\gamma}})$ with $\hat{w}_{ij}^1 = \tilde{w}_{ij}^1 \hat{x}_j$ and $\hat{w}_{ij}^2 = \tilde{w}_{ij}^2(1 - \hat{x}_j)$ for all(i, j), $i \neq j$, and with $\hat{\gamma}_{ij}^1 = \tilde{\gamma}_{ij}^1 \hat{x}_j$ and $\hat{\gamma}_{ij}^2 = \tilde{\gamma}_{ij}^2(1 - \hat{x}_j)$ for all (i, j) satisfies primal feasibility and complementary slackness to (25) and (26) by (18) and (21) since the inequalities are simply scaled by either the nonnegative value $\hat{x}_j$ or $1 - \hat{x}_j$. This completes the proof. □

The main result now follows.

Theorem 3. *The optimal objective function values to Problems NP and the continuous relaxation of QPRLT are equal, with any optimal set of dual values $\boldsymbol{\alpha}^1$, $\boldsymbol{\alpha}^2$, and $\boldsymbol{\alpha}^3$ to constraints* (27), (28), *and* (29) *of QPRLT, respectively, solving NP, where $\boldsymbol{\alpha} = (\boldsymbol{\alpha}^1, \boldsymbol{\alpha}^2, \boldsymbol{\alpha}^3)$.*

Proof. Since Problem LD is the Lagrangian dual to QPRLT obtained by placing constraints (27), (28), and (29) into the objective function using multipliers $\boldsymbol{\alpha} = (\boldsymbol{\alpha}^1, \boldsymbol{\alpha}^2, \boldsymbol{\alpha}^3)$, it follows directly that $\theta(\boldsymbol{\alpha})$ equals the optimal objective function value to the continuous relaxation of QPRLT at any $\boldsymbol{\alpha}$ comprising part of an optimal dual solution to this latter problem. By Theorem 2, this value in turn equals the optimal objective function value to $\text{LP}(\boldsymbol{\alpha})$ of (38). Now, suppose we delete the $4n$ inequalities (15) and (17), and the $2n$ righthand inequalities of (14) and (16) from $\text{G2}(\boldsymbol{\alpha})$. An optimal solution to the continuous relaxation of the resulting program must then have $z_j^1 = L_j^{\boldsymbol{\alpha}1} x_j$ and $z_j^2 = \bar{L}_j^{\boldsymbol{\alpha}0}(1 - x_j)$ for each j, providing the same objective value in the continuous relaxation of this reduced version of $\text{G2}(\boldsymbol{\alpha})$ as $\text{LP}(\boldsymbol{\alpha})$. Theorem 1 thus ensures that the optimal objective function value to the continuous relaxation of $\text{G2}(\boldsymbol{\alpha})$ must equal that of $\text{LP}(\boldsymbol{\alpha})$ at every such optimal $\boldsymbol{\alpha}$. This completes the proof. □

The above theorems and proofs collectively explain how to construct instances of $\text{G2}(\boldsymbol{\alpha})$ that provide the greatest possible relaxation value. Such constructions are based on optimal dual solutions to the continuous relaxation of QPRLT, permitting the optimal objective function values to the relaxations of $\text{G2}(\boldsymbol{\alpha})$ and QPRLT to equal. Given any such optimal dual solution, the $\boldsymbol{\alpha}$-vector used to compute $\text{G2}(\boldsymbol{\alpha})$ are the multipliers to (27), (28), and (29), respectively, as stated in Theorem 3. The decomposition argument in the proof of Theorem 2 essentially establishes the lefthand inequalities of (14) and (16) as surrogates of inequalities (25) and (26) using the prescribed optimal dual solutions. The proof of Theorem 1 demonstrates that all inequalities (14)–(17) are surrogates of inequalities (25)–(29). Hence, Problem $\text{G2}(\boldsymbol{\alpha})$ can be considered as a surrogate dual to a Lagrangian subproblem of QPRLT, where the equality restrictions (27), (28), and (29) are both dualized and treated as constraints.

Three remarks relative to G2($\boldsymbol{\alpha}$) are warranted. First, and as used in the proof of Theorem 3, Theorems 1 and 2 combine to show that for any dual-optimal $\boldsymbol{\alpha}$ to (27)–(29) of QPRLT, only the $2n$ lefthand inequalities of (14) and (16), together with the $(\boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{S}$ restrictions, are needed to have G2($\boldsymbol{\alpha}$) and QPRLT provide the same relaxation value. The additional $6n$ restrictions enforce that, for each $j = 1, \ldots, n$, $z_j^1 = g_j^{\boldsymbol{\alpha}}(\boldsymbol{x}, \boldsymbol{y})x_j$ and $z_j^2 = h_j^{\boldsymbol{\alpha}}(\boldsymbol{x}, \boldsymbol{y})(1 - x_j)$ at all $(\boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{X}$. This is in contrast to our discussion in Section 2.1.1 and Example 2.1 explaining that the omission of the righthand inequalities in (4) and (5) can alter the optimal relaxation value of G2. For general $\boldsymbol{\alpha}$, the righthand inequalities of (14)–(17) cannot be omitted in G2($\boldsymbol{\alpha}$) without potentially sacrificing relaxation strength, but such omissions can be performed for any dual-optimal $\boldsymbol{\alpha}$. Second, and as pointed out in Section 2.1 for Problem G, the $4n$ righthand inequalities in (14)–(17)are unnecessary in Problem G2($\boldsymbol{\alpha}$) since the desired equivalence between G2($\boldsymbol{\alpha}$) and QP is needed only at optimality. Finally, and again as noted in Section 2.1 for Problem G, a substitution of variables in terms of the slack variables for either (14) or (15), and in terms of the slack variables for either (16) or (17), will reduce the number of structural inequalities by $2n$. The net effect of the constraint eliminations and variable substitutions from the prior two remarks is to obtain an equivalent mixed 0-1 linear representation of QP that has only $2n$ auxiliary structural constraints in $2n$ additional nonnegative variables, and has the relaxation strength of the level-1 RLT formulation [23,24].

4. Exploiting special structure

Special structure in the constraints defining the set $\boldsymbol{S}$ of Problem QP can lead to more efficient implementations of [11] that give the level-1 RLT relaxation value. We consider two general structures. The first deals with instances where restrictions (28) and (29) in the relaxation of QPRLT all have multipliers of 0 in an optimal dual solution. Included within these instances is the family of quadratic set partitioning problems. The second arises when special subsets of the restrictions, fewer than $2n$, imply the bounding restrictions $\mathbf{0} \leqslant \boldsymbol{x} \leqslant \mathbf{1}$ so that a specially structured RLT [26] can be employed. For this second case, the relaxation strength of the specially structured RLT can exceed that of QPRLT.

4.1. Pure 0-1 programs with equality restrictions

Consider the implications of Theorem 3 when the relaxation of QPRLT is known to have an optimal dual solution with multipliers $\boldsymbol{\alpha}^2 = \mathbf{0}$ and $\boldsymbol{\alpha}^3 = \mathbf{0}$ corresponding to (28) and (29), respectively. The Theorem maintains that the optimal objective function values to Problems NP and the continuous relaxation of QPRLT equal, and asserts that $\boldsymbol{\alpha} = (\boldsymbol{\alpha}^1, \mathbf{0}, \mathbf{0})$ solves NP, where $\boldsymbol{\alpha}^1$ is any optimal set of dual values to (27). This is significant since, when such conditions are met, a linearization of QP having only n additional inequalities in n additional nonnegative variables with the strength of the level-1 RLT relaxation is possible. This is a savings of n inequality restrictions and n variables over the formulation of the previous section. The reason is that Problem G2($\boldsymbol{\alpha}$) will reduce in size. For such $\boldsymbol{\alpha}$ vectors, $h_j^{\boldsymbol{\alpha}}(\boldsymbol{x}, \boldsymbol{y}) = 0$ for all j by (13) so that programs (20) and (21) give $\bar{L}_j^{\boldsymbol{\alpha}1} = \bar{U}_j^{\boldsymbol{\alpha}1} = \bar{L}_j^{\boldsymbol{\alpha}0} = \bar{U}_j^{\boldsymbol{\alpha}0} = 0$ for all j. By (16) and (17), we then have that $z_j^2 = 0$ for all j in G2($\boldsymbol{\alpha}$). The formulation G2($\boldsymbol{\alpha}$) thus simplifies to G2$'$($\boldsymbol{\alpha}$) below, where we have recognized the righthand inequalities of (14) and (15) as redundant at optimality.

$$
\begin{aligned}
\text{G2}'(\boldsymbol{\alpha}) : \text{minimize} \quad & l^{\boldsymbol{\alpha}}(\boldsymbol{x}, \boldsymbol{y}) + \sum_{j=1}^{n} z_j^1 \\
\text{subject to} \quad & L_j^{\boldsymbol{\alpha}1} x_j \leqslant z_j^1 \quad \forall j && (39) \\
& g_j^{\boldsymbol{\alpha}}(\boldsymbol{x}, \boldsymbol{y}) - U_j^{\boldsymbol{\alpha}0}(1 - x_j) \leqslant z_j^1 \quad \forall j && (40) \\
& (\boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{X}
\end{aligned}
$$

As with Problems G2 and G2($\boldsymbol{\alpha}$), a substitution in terms of the slack variables to either set of constraints (39) or (40) can be made to obtain the desired formulation.

We now invoke the RLT theory to identify an important class of problems that have $\boldsymbol{\alpha}^2 = \mathbf{0}$ and $\boldsymbol{\alpha}^3 = \mathbf{0}$ in an optimal dual solution to the relaxation of QPRLT. Consider the special cases of QP where there are no continuous variables $\boldsymbol{y}$ and the constraints defining the set $\boldsymbol{S}$ are all equality, except for restrictions of the form $\boldsymbol{x} \geqslant \mathbf{0}$. Here, as before, $\boldsymbol{S}$ is assumed to imply $\boldsymbol{x} \leqslant \mathbf{1}$, though in this case such an assumption forfeits generality. Using obvious notation, Problem QP can be rewritten as QP$'$.

$$
\begin{aligned}
\text{QP}' : \text{minimize} \quad & l(\boldsymbol{x}) + \sum_{j=1}^{n} g_j(\boldsymbol{x}) x_j \\
\text{subject to} \quad & \boldsymbol{x} \in \boldsymbol{X} \equiv \{\boldsymbol{x} \in \boldsymbol{S} : \boldsymbol{x} \text{ binary}\}
\end{aligned}
$$

The set $\boldsymbol{S}$ and the linear functions $g_j(\boldsymbol{x})$ for all j simplify from their respective descriptions in (23) and (24) to the below.

$$\boldsymbol{S} = \left\{ \boldsymbol{x} \geqslant \boldsymbol{0} : \sum_{i=1}^{n} a_{ri} x_i = b_r \ \forall r = 1, \ldots, R \right\} \tag{41}$$

$$g_j(\boldsymbol{x}) = \sum_{\substack{i=1 \\ i \neq j}}^{n} C_{ij} x_i \quad \forall j = 1, \ldots, n \tag{42}$$

The RLT theory [23–25] does not require multiplying the equality restrictions of $\boldsymbol{S}$ by the factors $(1 - x_j)$ for all j provided that these equations of $\boldsymbol{S}$ are preserved in the level-1 representation. Nor does it require multiplying the $\boldsymbol{x} \geqslant \boldsymbol{0}$ inequalities by these same $(1 - x_j)$ factors. For both sets of multiplications, the resulting linearized inequalities would be implied by the other restrictions. The latter implication is due to $\boldsymbol{x} \in \boldsymbol{S}$ enforcing $\boldsymbol{x} \leqslant \boldsymbol{1}$, so that the set $\boldsymbol{S}$, together with the multiplication of the restrictions of $\boldsymbol{S}$ by the factors x_j, will imply such expressions. In other words, restrictions of the type (26) are redundant in the relaxation of QPRLT, making (28) and (29) also redundant so that $\boldsymbol{\alpha}^2 = \boldsymbol{0}$ and $\boldsymbol{\alpha}^3 = \boldsymbol{0}$ at an optimal dual solution as desired. (Each equation in (41) can be expressed as two inequalities to fit the form of (25).)

Observe that the family of quadratic set partitioning problems are encompassed by QP′ and therefore can be reformulated in terms of $\mathrm{G2}'(\boldsymbol{\alpha})$. Since $\mathrm{G2}'(\boldsymbol{\alpha})$ is a function of only the vector $\boldsymbol{\alpha}^1$ in such problems, a direct consequence of Theorem 3 is that the level-1 RLT relaxation value can be achieved by strategically "splitting" the objective coefficients on the quadratic terms $x_i x_j$ and $x_j x_i$ in such a manner that, for each (i, j) with $i < j$, the coefficient on the term $x_i x_j$ is decreased by the same quantity that the coefficient on the term $x_j x_i$ is increased. The vector $\boldsymbol{\alpha}^1$ dictates such a split by placing identities (8) into the objective function so that for each (i, j) with $i < j$, $x_i x_j$ is decreased and $x_j x_i$ is increased by the value α^1_{ij}.

Interestingly, as the celebrated quadratic assignment problem (QAP) is a quadratic set partitioning problem, it can be formulated in terms of $\mathrm{G2}'(\boldsymbol{\alpha})$. In fact, Kaufman and Broeckx [16] (see also [8,9] for related implementations) incorporated Problem G, less the redundant righthand inequalities (1) and (2), in both a mixed-integer solver and Benders' decomposition algorithm [6], but reported "disappointing" computational results. Hahn et al. [14], on the other hand, obtained superior results in an enumerative strategy that computes bounds obtainable from the level-1 RLT formulation (see [13]). These authors solved the Nugent et al. [22] size 25 test problem and the Krarup and Pruzan [18] size 30a problem to optimality. Our contention is that the performance difference between [14] and [16] is primarily due to the linearization strength. Adams and Johnson [2] showed the theoretical superiority of the level-1 RLT relaxation to the majority of published bounding strategies for the QAP. It appears promising, therefore, to combine the strength of the level-1 formulation with the conciseness of the linearization in [11] by suitably constructing $\mathrm{G2}'(\boldsymbol{\alpha})$. Of course, one must solve the level-1 relaxation to obtain the vector $\boldsymbol{\alpha}^1$, but the structure of QPRLT lends itself to efficient methods, as noted in [13,15]. Even so, an optimal $\boldsymbol{\alpha}^1$ is not required, a near-optimal dual solution suffices.

We conclude this section with an example to demonstrate the utility of splitting the objective coefficients for a quadratic set partitioning problem, and how the level-1 RLT relaxation provides such an optimal split.

Example 4.1. Consider an instance of Problem QP′ having $n = 7$ binary variables $\boldsymbol{x}$, where the function $l(\boldsymbol{x})$ is defined as

$$l(\boldsymbol{x}) = 6x_1 + 4x_2 + 5x_3 + 10x_4 + 6x_5 - 4x_6 + 3x_7,$$

where the coefficients C_{ij} of $g_j(\boldsymbol{x})$ found in (42) for each (i, j) are given by the (i, j)th entry of the matrix

$$\boldsymbol{C} = \begin{bmatrix} * & 0 & 0 & 0 & 0 & 0 & 0 \\ -10 & * & 0 & 0 & 0 & 0 & 0 \\ 8 & -6 & * & 0 & 0 & 0 & 0 \\ -4 & -8 & -6 & * & 0 & 0 & 0 \\ 0 & -8 & 10 & -10 & * & 0 & 0 \\ 1 & -10 & -3 & 0 & -8 & * & 0 \\ -10 & 0 & -6 & -10 & -6 & -8 & * \end{bmatrix},$$

and where the set $\boldsymbol{S}$ in (41) is defined as

$$\boldsymbol{S} = \{\boldsymbol{x} \geqslant \boldsymbol{0} : x_1 + x_2 + x_3 = 1, x_3 + x_4 + x_5 = 1, x_5 + x_6 + x_7 = 1\}.$$

The formulation of [11] with the bounds of Section 2.1.1, which is $\mathrm{G2}'(\boldsymbol{0})$, is obtained by computing $(L_1^{\boldsymbol{\alpha}1}, L_2^{\boldsymbol{\alpha}1}, L_3^{\boldsymbol{\alpha}1}, L_4^{\boldsymbol{\alpha}1}, L_5^{\boldsymbol{\alpha}1}, L_6^{\boldsymbol{\alpha}1}, L_7^{\boldsymbol{\alpha}1}) = (-14, -18, -6, -10, 0, 0, 0)$ and $(U_1^{\boldsymbol{\alpha}0}, U_2^{\boldsymbol{\alpha}0}, U_3^{\boldsymbol{\alpha}0}, U_4^{\boldsymbol{\alpha}0}, U_5^{\boldsymbol{\alpha}0}, U_6^{\boldsymbol{\alpha}0}, U_7^{\boldsymbol{\alpha}0}) = (9, -6, 10, 0, -6, 0, 0)$ as prescribed in the minimization problems of (18) and the maximization problems of (19), respectively, to generate (39) and (40). The optimal objective value to the continuous relaxation of G2′($\mathbf{0}$) is -10.50.

Next construct QPRLT. Following the discussion of this section, inequalities (26), (28), and (29) are not necessary. Furthermore, constraints (25) are equality, and the coefficients D_{ij} and d_{ri} are all 0 as there are no continuous variables. The optimal objective value to the continuous relaxation of QPRLT is -8, which is the integer optimum objective.

Upon solving this relaxation of QPRLT, we obtain an optimal set of dual variables $\hat{\boldsymbol{\alpha}}^1$ to constraints (27) with $\hat{\alpha}^1_{ij}$ for each (i, j), $i < j$, given by the (i, j)th entry of the matrix.

$$\hat{\boldsymbol{\alpha}}^1 = \begin{bmatrix} * & 0 & 0 & 4 & 16 & -5 & 6 \\ * & * & 0 & 7 & 6 & 9 & 6 \\ * & * & * & 0 & 0 & 3 & 6 \\ * & * & * & * & 0 & 0 & 10 \\ * & * & * & * & * & 0 & 0 \\ * & * & * & * & * & * & 0 \\ * & * & * & * & * & * & * \end{bmatrix}$$

We then "split" the quadratic objective coefficients by setting $C'_{ij} = C_{ij} - \hat{\alpha}^1_{ij}$ and $C'_{ji} = C_{ji} + \hat{\alpha}^1_{ij}$ for all $i < j$ to obtain the following new quadratic cost matrix $\boldsymbol{C}'$:

$$\boldsymbol{C}' = \begin{bmatrix} * & 0 & 0 & -4 & -16 & 5 & -6 \\ -10 & * & 0 & -7 & -6 & -9 & -6 \\ 8 & -6 & * & 0 & 0 & -3 & -6 \\ 0 & -1 & -6 & * & 0 & 0 & -10 \\ 16 & -2 & 10 & -10 & * & 0 & 0 \\ -4 & -1 & 0 & 0 & -8 & * & 0 \\ -4 & 6 & 0 & 0 & -6 & -8 & * \end{bmatrix}$$

Problem G2′($\hat{\boldsymbol{\alpha}}$) with $\hat{\boldsymbol{\alpha}} = (\hat{\boldsymbol{\alpha}}^1, \mathbf{0}, \mathbf{0})$ is the formulation of [11] with strengthened bounds and cost matrix $\boldsymbol{C}'$. This formulation is obtained by computing $(L_1^{\boldsymbol{\alpha}1}, L_2^{\boldsymbol{\alpha}1}, L_3^{\boldsymbol{\alpha}1}, L_4^{\boldsymbol{\alpha}1}, L_5^{\boldsymbol{\alpha}1}, L_6^{\boldsymbol{\alpha}1}, L_7^{\boldsymbol{\alpha}1}) = (-4, -2, 0, -7, -16, -9, -16)$ and $(U_1^{\boldsymbol{\alpha}0}, U_2^{\boldsymbol{\alpha}0}, U_3^{\boldsymbol{\alpha}0}, U_4^{\boldsymbol{\alpha}0}, U_5^{\boldsymbol{\alpha}0}, U_6^{\boldsymbol{\alpha}0}, U_7^{\boldsymbol{\alpha}0}) = (6, 5, 10, 0, -6, 5, -6)$. The continuous relaxation has an optimal objective function value of -8, with an optimal (binary) solution given by $(x_1, x_2, x_3, x_4, x_5, x_6, x_7) = (0, 1, 0, 1, 0, 1, 0)$. Consistent with the result of Theorem 3 and the discussion of this section, this is the bound yielded by the level-1 RLT.

4.2. Structured binary functions

We consider in this section a special case of Problem QP where the restrictions defining the set $\boldsymbol{X}$ give rise to p linear functions $f_k(\boldsymbol{x})$, $k = 1, \ldots, p$, of the binary variables $\boldsymbol{x}$ that satisfy the following two conditions:

1. each linear function $f_k(\boldsymbol{x})$ for $k = 1, \ldots, p$ realizes either the value 0 or 1 at every $(\boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{X}$, and
2. the (valid) linear inequalities $f_k(\boldsymbol{x}) \geqslant 0$ for $k = 1, \ldots, p$ imply that:
   (a) $\mathbf{0} \leqslant \boldsymbol{x} \leqslant \mathbf{1}$, and
   (b) $f_k(\boldsymbol{x}) \leqslant 1$ for $k = 1, \ldots, p$.

We assume without loss of generality that the inequalities $f_k(\boldsymbol{x}) \geqslant 0$ do not imply for any j that either $x_j > 0$ or $x_j < 1$ since otherwise the variable x_j can be fixed to a binary value and removed from the problem. We also henceforth assume for each $k = 1, \ldots, p$ that $\min\{f_k(\boldsymbol{x}) : \boldsymbol{x} \in \boldsymbol{S}\} = 0$ and $\max\{f_k(\boldsymbol{x}) : \boldsymbol{x} \in \boldsymbol{S}\} = 1$ since otherwise $f_k(\boldsymbol{x})$ can be fixed to a binary value.

Such functions $f_k(\boldsymbol{x})$ may be explicitly found in QP, or can result from substitutions and/or scalings. For example, given an equation $\sum_{i=1}^n a_{ri} x_i = b_r$ in $\boldsymbol{S}$ where a_{rn} is nonzero, the expression $(b_r - \sum_{i=1}^{n-1} a_{ri} x_i)/a_{rn}$ equals the binary variable x_n, so that this expression can serve as such a function $f_k(\boldsymbol{x})$. In order to exploit these functions to obtain more concise representations than QPRLT, however, we need $p < 2n$. As we will later discuss, such a collection of $p < 2n$ restrictions satisfying conditions 1 and 2 arise from various special structures, including variable upper bounding and generalized upper bounding. For convenience, we represent the functions $f_k(\boldsymbol{x})$ as

$$f_k(\boldsymbol{x}) = \beta_{k0} + \beta_{k1} x_1 + \beta_{k2} x_2 + \cdots + \beta_{kn} x_n \quad \forall k = 1, \ldots, p \tag{43}$$

where β_{kj} for all (k, j), $k = 1, \ldots, p$, $j = 0, \ldots, n$, are scalars.

We consider a special application of the method of Glover [11] that exploits such functions $f_k(\boldsymbol{x})$ in the following section.

4.2.1. Glover's method

Given $f_k(\boldsymbol{x})$, $k = 1, \ldots, p$, satisfying conditions 1 and 2, suppose we rewrite the objective function to Problem QP in terms of these functions and $p+1$ additional functions $\psi_k^{\boldsymbol{\pi}}(\boldsymbol{x}, \boldsymbol{y})$, $k = 0, \ldots, p$, so that

$$\sum_{j=1}^{n} g_j(\boldsymbol{x}, \boldsymbol{y}) x_j = \sum_{k=1}^{p} \psi_k^{\boldsymbol{\pi}}(\boldsymbol{x}, \boldsymbol{y}) f_k(\boldsymbol{x}) - \psi_0^{\boldsymbol{\pi}}(\boldsymbol{x}, \boldsymbol{y}) \quad \text{for all } (\boldsymbol{x}, \boldsymbol{y}), \boldsymbol{x} \text{ binary.} \tag{44}$$

Here, for each $k = 1, \ldots, p$, the expression $\psi_k^{\boldsymbol{\pi}}(\boldsymbol{x}, \boldsymbol{y})$ is a linear function of the variables $\boldsymbol{x}$ and $\boldsymbol{y}$ whose coefficients are defined in terms of the vector $\boldsymbol{\pi} = (\boldsymbol{\pi}^1, \boldsymbol{\pi}^2)$ as

$$\psi_k^{\boldsymbol{\pi}}(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{n} \pi_{ik}^1 x_i + \sum_{i=1}^{m} \pi_{ik}^2 y_i. \tag{45}$$

The function $\psi_0^{\boldsymbol{\pi}}(\boldsymbol{x}, \boldsymbol{y})$ is also linear in the variables $\boldsymbol{x}$ and $\boldsymbol{y}$, and must be defined in terms of the vector $\boldsymbol{\pi}$ as

$$\psi_0^{\boldsymbol{\pi}}(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{n} \left[\sum_{k=1}^{p} \pi_{ik}^1 (\beta_{k0} + \beta_{ki}) \right] x_i + \sum_{i=1}^{m} \left(\sum_{k=1}^{p} \pi_{ik}^2 \beta_{k0} \right) y_i \tag{46}$$

for (44) to hold true. Recall that we had earlier assumed without loss of generality that for each j, the expression $g_j(\boldsymbol{x}, \boldsymbol{y})$ is not a function of the variable x_j and that it does not contain a term of degree 0. Hence, the lefthand sum in (44) has no linear terms. The function $\psi_0^{\boldsymbol{\pi}}(\boldsymbol{x}, \boldsymbol{y})$ compensates for the linear terms within the products $\psi_k^{\boldsymbol{\pi}}(\boldsymbol{x}, \boldsymbol{y}) f_k(\boldsymbol{x})$ of the righthand sum in (44), including the x_i variables arising from the substitution $x_i = x_i^2$.

Functions $\psi_k^{\boldsymbol{\pi}}(\boldsymbol{x}, \boldsymbol{y})$ for $k = 0, \ldots, p$ satisfying (44) must exist, and they are not necessarily unique. Observe for each j that since the linear inequalities $f_k(\boldsymbol{x}) \geqslant 0$ for $k = 1, \ldots, p$ imply $x_j \geqslant 0$ by condition 2a, there must exist a nonnegative linear combination of the functions $f_k(\boldsymbol{x})$ with multipliers, say ω_k, yielding x_j: that is, $\sum_{k=1}^{p} \omega_k f_k(\boldsymbol{x}) = x_j$. Consequently, $g_j(\boldsymbol{x}, \boldsymbol{y}) x_j$ can be expressed as $\sum_{k=1}^{p} [g_j(\boldsymbol{x}, \boldsymbol{y}) \omega_k] f_k(\boldsymbol{x})$ so that $\psi_k^{\boldsymbol{\pi}}(\boldsymbol{x}, \boldsymbol{y}) = g_j(\boldsymbol{x}, \boldsymbol{y}) \omega_k$ for all $k = 1, \ldots, p$ and $\psi_0^{\boldsymbol{\pi}}(\boldsymbol{x}, \boldsymbol{y}) = 0$ for this special case. We can therefore sequentially progress through each j and adjust the functions $\psi_k^{\boldsymbol{\pi}}(\boldsymbol{x}, \boldsymbol{y})$ accordingly to satisfy (44).

Now, given a vector $\boldsymbol{\pi}$ for which (44) holds true, the idea is to linearize Problem QP by substituting for each $k = 1, \ldots, p$, a continuous variable z_k for the product $\psi_k^{\boldsymbol{\pi}}(\boldsymbol{x}, \boldsymbol{y}) f_k(\boldsymbol{x})$ in the objective function. In the same spirit as (39) and (40), we will then devise linear restrictions that ensure $z_k = \psi_k^{\boldsymbol{\pi}}(\boldsymbol{x}, \boldsymbol{y}) f_k(\boldsymbol{x})$ for each $k = 1, \ldots, p$ at optimality to the linear problem. Toward this end, compute for each $k = 1, \ldots, p$, values $\mathscr{L}_k^{\boldsymbol{\pi}}$ and $\mathscr{U}_k^{\boldsymbol{\pi}}$ as

$$\begin{aligned} \mathscr{L}_k^{\boldsymbol{\pi}} &= \min\{\psi_k^{\boldsymbol{\pi}}(\boldsymbol{x}, \boldsymbol{y}) : (\boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{X}^R,\ f_k(\boldsymbol{x}) = 1\} \quad \text{and} \\ \mathscr{U}_k^{\boldsymbol{\pi}} &= \max\{\psi_k^{\boldsymbol{\pi}}(\boldsymbol{x}, \boldsymbol{y}) : (\boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{X}^R,\ f_k(\boldsymbol{x}) = 0\} \end{aligned} \tag{47}$$

where, as with (3), these programs are assumed bounded. Consistent with the definition in Section 2, the set $\boldsymbol{X}^R$ is any relaxation of $\boldsymbol{X}$ in the variables $(\boldsymbol{x}, \boldsymbol{y})$. Then form the following program.

$$\begin{aligned} \text{G3}(\boldsymbol{\pi}) : \text{minimize} \quad & l(\boldsymbol{x}, \boldsymbol{y}) - \psi_0^{\boldsymbol{\pi}}(\boldsymbol{x}, \boldsymbol{y}) + \sum_{k=1}^{p} z_k \\ \text{subject to} \quad & \mathscr{L}_k^{\boldsymbol{\pi}} f_k(\boldsymbol{x}) \leqslant z_k \quad \forall k = 1, \ldots, p \qquad (48) \\ & \psi_k^{\boldsymbol{\pi}}(\boldsymbol{x}, \boldsymbol{y}) - \mathscr{U}_k^{\boldsymbol{\pi}} (1 - f_k(\boldsymbol{x})) \leqslant z_k \quad \forall k = 1, \ldots, p \qquad (49) \\ & (\boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{X} \end{aligned}$$

Problem G3($\boldsymbol{\pi}$) is our equivalent mixed 0-1 linear reformulation of QP. As desired, inequalities (48) and (49) enforce for each $k = 1, \ldots, p$ that $z_k = \psi_k^{\boldsymbol{\pi}}(\boldsymbol{x}, \boldsymbol{y}) f_k(\boldsymbol{x})$ at optimality. To see this, consider any $k = 1, \ldots, p$, and any $(\boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{X}$. Condition 1 stipulates that $f_k(\boldsymbol{x})$ equals either 0 or 1. If $f_k(\boldsymbol{x}) = 0$, inequalities (48) enforce $z_k = 0$ at optimality with (49) redundant. If $f_k(\boldsymbol{x}) = 1$, inequalities (49) enforce $z_k = \psi_k^{\boldsymbol{\pi}}(\boldsymbol{x}, \boldsymbol{y})$ at optimality with (48) redundant. Hence, $z_k = \psi_k^{\boldsymbol{\pi}}(\boldsymbol{x}, \boldsymbol{y}) f_k(\boldsymbol{x})$.

Observe that the construction of G3($\boldsymbol{\pi}$) requires only that the functions $f_k(\boldsymbol{x})$ satisfy condition 1 and the $\boldsymbol{x} \geqslant \boldsymbol{0}$ restrictions of 2a. Condition 1 ensures that $z_k = \psi_k^{\boldsymbol{\pi}}(\boldsymbol{x}, \boldsymbol{y}) f_k(\boldsymbol{x})$ for all $k = 1, \ldots, p$ at optimality to G3($\boldsymbol{\pi}$) while the nonnegativity restrictions on $\boldsymbol{x}$ of condition 2a establish the existence of functions $\psi_k^{\boldsymbol{\pi}}(\boldsymbol{x},\boldsymbol{y})$, $k=0,\ldots,p$, satisfying (44). The $\boldsymbol{x}\leqslant \mathbf{1}$ restrictions of 2a and condition 2b are not used here, but are needed in the construction of the special-structure RLT in the upcoming section.

G3($\boldsymbol{\pi}$) compares favorably to G2($\boldsymbol{\alpha}$) in terms of problem size when $p<2n$. Recall from the concluding paragraph of Section 3 that a reduced version of G2($\boldsymbol{\alpha}$) requires $2n$ auxiliary variables and $2n$ additional structural constraints, since the righthand restrictions in (14)–(17) are not necessary to ensure equivalence between Problems G2($\boldsymbol{\alpha}$) and QP at optimality, and since substitutions in terms of slack variables can be made. G3($\boldsymbol{\pi}$), on the other hand, involves p additional variables and $2p$ additional constraints. As with G2($\boldsymbol{\alpha}$), the size of G3($\boldsymbol{\pi}$) can be reduced by performing a substitution of variables in terms of the slacks for either (48) or (49), resulting in only p auxiliary constraints in p additional nonnegative variables z_k. This is a savings of $2n-p$ variables and $2n-p$ constraints realized by the special-structure formulation over the standard model.

Depending on the vector $\boldsymbol{\pi}$, the continuous relaxation of G3($\boldsymbol{\pi}$) obtained by relaxing the $(\boldsymbol{x},\boldsymbol{y})\in \boldsymbol{X}$ restrictions to $(\boldsymbol{x},\boldsymbol{y})\in \boldsymbol{S}$ can have different optimal objective function values. It is desired to obtain a vector $\boldsymbol{\pi}$ that satisfies (44) and yields the maximum such objective value. Notationally, we wish to solve the nonlinear (special-structure) problem:

$$\text{NSP}: \eta^* = \max_{\boldsymbol{\pi} \text{ satisfies } (44)} \eta(\boldsymbol{\pi}) = \min \left\{ l(\boldsymbol{x},\boldsymbol{y}) - \psi_0^{\boldsymbol{\pi}}(\boldsymbol{x},\boldsymbol{y}) + \sum_{k=1}^{p} z_k : (48), (49), (\boldsymbol{x},\boldsymbol{y})\in \boldsymbol{S} \right\}. \tag{50}$$

We consider in the following section an application of the RLT using functions $f_k(\boldsymbol{x})$ that satisfy the prescribed conditions 1 and 2.

4.2.2. *Special-structure RLT*

Given p linear functions $f_k(\boldsymbol{x})$, $k=1,\ldots,p$, satisfying conditions 1 and 2, the special-structure RLT theory of [26] motivates a linear reformulation of QP that has a relaxation strength at least that of QPRLT. The key ingredient is that the nonnegative functions $f_k(\boldsymbol{x})$ for $k=1,\ldots,p$ are used as the product factors in lieu of the standard factors x_j and $(1-x_j)$ for all $j=1,\ldots,n$. The idea is that, since the nonnegativity of these special $f_k(\boldsymbol{x})$ factors implies the nonnegativity of the standard factors as set forth in condition 2a, the linearization resulting from these special factors will also imply the standard linearization.

The derivation of the special-structure level-1 RLT linearization proceeds in a similar manner to the construction of QPRLT. The reformulation step multiplies every constraint in (23) defining the set $\boldsymbol{S}$ by each $f_k(\boldsymbol{x})$, and appends these Rp new restrictions to QP, substituting throughout the binary identity that $x_j^2=x_j$ for all j. Here, we choose in the linearization step to substitute a continuous variable w_{ij} for every occurrence of *either* product x_ix_j or x_jx_i for all (i,j), $i<j$, and a continuous variable γ_{ij} for every occurrence of the product x_iy_j (equivalently y_jx_i) for all (i,j), $i=1,\ldots,m$, within the objective function and constraints. We let the notation $\lfloor x_i f_k(\boldsymbol{x})\rfloor_L$, $\lfloor y_i f_k(\boldsymbol{x})\rfloor_L$, and $\lfloor g_j(\boldsymbol{x},\boldsymbol{y})x_j\rfloor_L$ denote the linearized versions of $x_i f_k(\boldsymbol{x})$, $y_i f_k(\boldsymbol{x})$, and $g_j(\boldsymbol{x},\boldsymbol{y})x_j$, respectively, under such substitutions. We then introduce continuous variables $v_i^k=\lfloor x_i f_k(\boldsymbol{x})\rfloor_L$ $\forall\ (i,k)$, $k=1,\ldots,p$, and $\lambda_i^k=\lfloor y_i f_k(\boldsymbol{x})\rfloor_L$ $\forall\ (i,k)$, $i=1,\ldots,m$, $k=1,\ldots,p$. These variables are substituted throughout the Rp new inequalities, with $p(n+m)$ constraints used to explicitly equate these variables to their substituted expressions. Finally, since condition 1 ensures for each k that $f_k(\boldsymbol{x})f_k(\boldsymbol{x})=f_k(\boldsymbol{x})$ for all $(\boldsymbol{x},\boldsymbol{y})\in \boldsymbol{X}$, we explicitly enforce that $\lfloor f_k(\boldsymbol{x})f_k(\boldsymbol{x})\rfloor_L=f_k(\boldsymbol{x})$ for all $k=1,\ldots,p$. Problem SQPRLT, the version of QPRLT resulting from this application of the special-structure RLT, is as follows.

$$\text{SQPRLT}: \text{minimize} \quad l(\boldsymbol{x},\boldsymbol{y}) + \sum_{j=1}^{n} \lfloor g_j(\boldsymbol{x},\boldsymbol{y})x_j\rfloor_L$$

$$\text{subject to} \quad \sum_{i=1}^{n} a_{ri}v_i^k + \sum_{i=1}^{m} d_{ri}\lambda_i^k \geqslant b_r f_k(\boldsymbol{x}) \quad \forall (r,k),\ r=1,\ldots,R,\ k=1,\ldots,p \tag{51}$$

$$\sum_{i=1}^{n} \beta_{ki}v_i^k = (1-\beta_{k0})f_k(\boldsymbol{x}) \quad \forall k=1,\ldots,p \tag{52}$$

$$\lfloor x_i f_k(\boldsymbol{x})\rfloor_L - v_i^k = 0 \quad \forall (i,k), k=1,\ldots,p \tag{53}$$

$$\lfloor y_i f_k(\boldsymbol{x})\rfloor_L - \lambda_i^k = 0 \quad \forall (i,k),\ i=1,\ldots,m,\ k=1,\ldots,p \tag{54}$$

$$(\boldsymbol{x},\boldsymbol{y})\in \boldsymbol{S} \tag{55}$$

$$\boldsymbol{x} \text{ binary}$$

Problems QPRLT and SQPRLT are similar in structure. Inequalities (51) are of the same type as (25) and (26), while (53) and (54) are of the form (27)–(29). The w_{ij}^1 and w_{ij}^2 variables of QPRLT are absorbed in the v_i^k variables of SQPRLT, as are the γ_{ij}^1

and γ^2_{ij} variables absorbed in the λ^k_i. The variables w^1_{ij} and γ^1_{ij} of QPRLT essentially play the role of w_{ij} and γ_{ij}, respectively, in the objective function and restrictions (53) and (54) of SQPRLT. Eqs. (52) are the restrictions $\lfloor f_k(\boldsymbol{x}) f_k(\boldsymbol{x})\rfloor_L = f_k(\boldsymbol{x})$ for all $k = 1, \ldots, p$, upon noting from (43) that $\lfloor f_k(\boldsymbol{x}) f_k(\boldsymbol{x})\rfloor_L = \beta_{k0} f_k(\boldsymbol{x}) + \sum_{i=1}^n \beta_{ki} v^k_i$ for each $k = 1, \ldots, p$. Observe that, in fact, SQPRLT reduces to QPRLT when the $f_k(\boldsymbol{x})$ factors default tox_j and $(1 - x_j)$ for all j. Similar to (30) of QPRLT, restrictions (55) enforcing $(\boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{S}$ are not necessary in SQPRLT as they are implied by (51), but we maintain them for convenience.

A final comment relative to the construction of SQPRLT will be used in the next section. We do not multiply the constraints in (23) by each of the p nonnegative expressions $1 - f_k(\boldsymbol{x})$ to create the Rp additional restrictions

$$\sum_{i=1}^n a_{ri}(x_i - v^k_i) + \sum_{i=1}^m d_{ri}(y_i - \lambda^k_i) \geqslant b_r(1 - f_k(\boldsymbol{x})) \quad \forall (r, k),\ r = 1, \ldots, R,\ k = 1, \ldots, p. \tag{56}$$

Condition 2b has that the inequalities $f_k(\boldsymbol{x}) \geqslant 0$ for all $k = 1, \ldots, p$ collectively imply $1 - f_k(\boldsymbol{x}) \geqslant 0$ for each k, so that the RLT theory [26] assures inequalities (56) are implied by (51) in the continuous relaxation of SQPRLT. This was our reason for originally introducing condition 2b.

4.2.3. *Enhancing Glover's method with the RLT*

The structure of Problem QPRLT that permitted its reformulation as a concise mixed 0-1 linear program having the strength of the level-1 RLT relaxation is also found in SQPRLT, so that the arguments of Section 3 found in Theorems 1–3 carry over directly to the special-structure instance. For such cases, this leads to a concise formulation of the size of G3($\boldsymbol{\pi}$) having the same relaxation strength as SQPRLT. Similar to our arguments in Section 3, we assume that the set $\boldsymbol{X}^R$ used to compute the bounds (47), found in (48) and (49) of G3($\boldsymbol{\pi}$), has $\boldsymbol{X}^R = \boldsymbol{S}$.

Consider first the relationship between the optimal objective function values to the continuous relaxations of Problems SQPRLT and G3($\boldsymbol{\pi}$). The theorem and proof below show that the former value is at least as large as the latter.

Theorem 4. *Given any vector* $\boldsymbol{\pi}$ *satisfying* (44), *the optimal objective function value to the continuous relaxation of Problem SQPRLT is an upper bound on the optimal objective value to the relaxation of* $G3(\boldsymbol{\pi})$.

Proof. Consider any vector $\boldsymbol{\pi}$ satisfying (44). Using obvious vector notation, it suffices to show that, given any feasible solution $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{w}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{v}}, \hat{\boldsymbol{\lambda}})$ to the continuous relaxation of SQPRLT, the point $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}})$ having $\hat{z}_k = \lfloor \psi^{\boldsymbol{\pi}}_k(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) f_k(\hat{\boldsymbol{x}})\rfloor_L$ for all k is feasible to the relaxation of G3($\boldsymbol{\pi}$) with the same objective function value. For each $k = 1, \ldots, p$, surrogate inequalities (51) and Eq. (52) with an optimal set of dual multipliers to the $(\boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{S}$ restrictions and the $f_k(\boldsymbol{x}) = 1$ constraint of the minimization problem in (47), respectively, to verify by (53) and (54) that $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}})$ satisfies (48). Now, recalling from condition 2b that inequalities (56) are implied by (51), surrogate (56) and (52) with a computed optimal set of dual multipliers to the $(\boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{S}$ restrictions and the negative of the computed dual value to the $f_k(\boldsymbol{x}) = 1$ constraint of the maximization problem in (47), respectively, to verify by (53) and (54) that $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}})$ satisfies (49). Hence $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}})$ is feasible to the continuous relaxation of G3($\boldsymbol{\pi}$). The objective function value to G3($\boldsymbol{\pi}$) at this point is $l(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) - \psi^{\boldsymbol{\pi}}_0(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) + \sum_{k=1}^p \hat{z}_k$, which equals the objective value to the relaxation of SQPRLT at $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{w}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{v}}, \hat{\boldsymbol{\lambda}})$ by (44). This completes the proof. □

The block diagonal structure of QPRLT demonstrated in the proof of Theorem 2 is present in SQPRLT. To see this, suppose we form a Lagrangian dual to the continuous relaxation of SQPRLT by placing constraints (53) and (54) into the objective function. Then the subproblem over (51), (52), and (55) can be solved via $p + 1$ independent blocks. Specifically, consider such a Lagrangian dual where we use the same vector notation $\boldsymbol{\pi} = (\boldsymbol{\pi}^1, \boldsymbol{\pi}^2)$ for the dual multipliers to (53) and (54) as was used in equations (44)–(46) of Section 4.2.1.

$$\text{SLD}: \quad \text{maximize } \tau(\boldsymbol{\pi})$$

where

$$\begin{aligned}\tau(\boldsymbol{\pi}) = \min \Bigg\{ & l(\boldsymbol{x}, \boldsymbol{y}) + \sum_{j=1}^n \lfloor g_j(\boldsymbol{x}, \boldsymbol{y}) x_j \rfloor_L \\ & + \sum_{k=1}^p \left[\sum_{i=1}^n \pi^1_{ik}(v^k_i - \lfloor x_i f_k(\boldsymbol{x})\rfloor_L) + \sum_{i=1}^m \pi^2_{ik}(\lambda^k_i - \lfloor y_i f_k(\boldsymbol{x})\rfloor_L) \right] : (51), (52), (55) \Bigg\}\end{aligned} \tag{57}$$

We restrict attention to those instances of $\boldsymbol{\pi}$ that permit a dual feasible completion to the relaxation of SQPRLT (so that dual feasibility with respect to the w_{ij} and γ_{ij} variables is satisfied). Otherwise, the Lagrangian subproblem $\tau(\boldsymbol{\pi})$ is unbounded below since (51), (52), and (55) of (57) do not involve the variables w_{ij} or γ_{ij}. Observe that this restriction is precisely the same as enforcing that the vector $\boldsymbol{\pi}$ satisfies (44). Consequently, for such values of $\boldsymbol{\pi}$, the calculation of $\tau(\boldsymbol{\pi})$ simplifies to:

$$\tau(\boldsymbol{\pi}) = \min \left\{ l(\boldsymbol{x}, \boldsymbol{y}) - \psi_0^{\boldsymbol{\pi}}(\boldsymbol{x}, \boldsymbol{y}) + \sum_{k=1}^{p} \left[\sum_{i=1}^{n} \pi_{ik}^1 v_i^k + \sum_{i=1}^{m} \pi_{ik}^2 \lambda_i^k \right] : (51), (52), (55) \right\}. \tag{58}$$

The block diagonal structure of the Lagrangian dual problem (57), equivalently (58), is exposed in the proof of the following theorem. This theorem and proof, designed for the special-structure linearization SQPRLT, parallels Theorem 2 and its proof.

Theorem 5. *Given any vector $\boldsymbol{\pi}$ satisfying* (44), *the value $\tau(\boldsymbol{\pi})$ in* (58) *is equal to the optimal objective function value of the special-structure linear program*

$$\text{SLP } (\boldsymbol{\pi}) : \text{minimize} \left\{ l(\boldsymbol{x}, \boldsymbol{y}) - \psi_0^{\boldsymbol{\pi}}(\boldsymbol{x}, \boldsymbol{y}) + \sum_{k=1}^{p} \mathscr{L}_k^{\boldsymbol{\pi}} f_k(\boldsymbol{x}) : (\boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{S} \right\}, \tag{59}$$

where for each k, $\mathscr{L}_k^{\boldsymbol{\pi}}$ is computed as in (47).

Proof. It is sufficient to show for each k that an optimal set of dual multipliers to the corresponding constraints in (51) and (52) of (58) can be computed using any optimal dual solution to the minimization problem in (47). The reason for this is that the dual to SLP($\boldsymbol{\pi}$) must then be the dual to $\tau(\boldsymbol{\pi})$ of (58), where the multipliers to (51) and (52) of $\tau(\boldsymbol{\pi})$ in (58) have been fixed in SLP($\boldsymbol{\pi}$) at an optimal set of values.

Given any $k = 1, \ldots, p$, solve the minimization problem in (47) to obtain a primal optimal solution $\tilde{v}_i^k$ for all $i = 1, \ldots, n$ and $\tilde{\lambda}_i^k$ for all $i = 1, \ldots, m$, to represent the x_i and y_i variables, respectively. Fix the dual multipliers to the associated inequality restrictions in (51) of (58) to the computed optimal duals to the constraints of $\boldsymbol{S}$ in (47), and fix the dual to the associated equation in (52) of (58) to the computed dual to the $f_k(\boldsymbol{x}) = 1$ restriction. Progress through each $k = 1, \ldots, p$, to obtain multipliers for all restrictions in (51) and (52). Solve the dual to (58) with these dual values fixed to obtain an $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) \in \boldsymbol{S}$ and multipliers $\boldsymbol{\xi}$. The fixed duals to (51) and (52), together with $\boldsymbol{\xi}$, define a dual feasible solution to (58), and$(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ and $\boldsymbol{\xi}$ satisfy complementary slackness to (55). Finally, $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{v}}, \hat{\boldsymbol{\lambda}})$ with $\hat{v}_i^k = \tilde{v}_i^k f_k(\hat{\boldsymbol{x}})$ for all (i, k), $k = 1, \ldots, p$, and $\hat{\lambda}_i^k = \tilde{\lambda}_i^k f_k(\hat{\boldsymbol{x}})$ for all (i, k), $i = 1, \ldots, m$, and $k = 1, \ldots, p$, satisfies primal feasibility and complementary slackness to (51) and (52) by (47) since restrictions (51) and (52) are scaled by the nonnegative $f_k(\hat{\boldsymbol{x}})$. This completes the proof. □

The relationship between the objective function values to Problems SQPRLT and G3($\boldsymbol{\pi}$) for vectors $\boldsymbol{\pi}$ satisfying (44) is considered in the following theorem. This theorem parallels Theorem 3 for Problems SQPRLT and G2($\boldsymbol{\alpha}$).

Theorem 6. *The optimal objective function values to Problems NSP and the continuous relaxation of SQPRLT are equal, with any optimal set of dual values $\boldsymbol{\pi}^1$ and $\boldsymbol{\pi}^2$ to constraints* (53) *and* (54) *of SQPRLT, respectively, solving NSP, where $\boldsymbol{\pi} = (\boldsymbol{\pi}^1, \boldsymbol{\pi}^2)$.*

Proof. Problem SLD is the Lagrangian dual of the continuous relaxation of SQPRLT obtained by placing equations (53) and (54) into the objective function using multipliers $\boldsymbol{\pi} = (\boldsymbol{\pi}^1, \boldsymbol{\pi}^2)$. Consequently, $\tau(\boldsymbol{\pi})$ given in (57) equals the optimal objective value to the continuous relaxation of SQPRLT at any vector $\boldsymbol{\pi}$ constituting part of an optimal dual solution. But since the variables $\boldsymbol{w}$ and $\boldsymbol{\gamma}$ appear only in constraints (53) and (54) of SQPRLT, such an optimal $\boldsymbol{\pi}$ must satisfy (44), so that $\tau(\boldsymbol{\pi})$ simplifies from (57) to (58). Theorem 5 states that $\tau(\boldsymbol{\pi})$ is equal to the optimal objective value to SLP($\boldsymbol{\pi}$). Now consider Problem G3($\boldsymbol{\pi}$) without the p inequalities (49). An optimal solution to the continuous relaxation of this reduced problem must have $z_k = \mathscr{L}_k^{\boldsymbol{\pi}} f_k(\boldsymbol{x})$ for each $k = 1, \ldots p$, yielding the same optimal objective value as SLP($\boldsymbol{\pi}$). Theorem 4 then gives that the optimal objective value to the continuous relaxation of G3($\boldsymbol{\pi}$) must equal that of $\tau(\boldsymbol{\pi})$ at every such optimal $\boldsymbol{\pi}$.This completes the proof. □

The net effect of Theorems 4–6 is to establish, for instances of QP promoting functions $f_k(\boldsymbol{x})$ that satisfy the prescribed conditions 1 and 2, concise linear reformulations of the form G3($\boldsymbol{\pi}$) that have tight continuous relaxations. Not only are the formulations G3($\boldsymbol{\pi}$) more concise than G2($\boldsymbol{\alpha}$) when $p < 2n$, but they can also promote tighter continuous relaxations. Recalling from the discussion at the beginning of Section 4.2.2 that the continuous relaxation of SQPRLT is at least as tight as that of QPRLT, it follows from Theorems 4 and 6 that $\eta^* \geqslant v^*$, where η^* and v^* are as defined in (50) and (22), respectively. Moreover, the formulation SQPRLT will also have fewer variables and constraints than QPRLT when $p < 2n$, which can affect the effort required to optimally solve Problems NSP and NP.

Functions $f_k(\boldsymbol{x})$ for $k=1,\ldots,p$ having $p<2n$ and satisfying conditions 1 and 2 arise in practice. One instance is variable upper bounding where certain constraints have the form $x_i \leqslant x_j$. In particular, consider an instance of Problem QP where there exists a subset of the n binary variables $\boldsymbol{x}$, say $x_1, x_2, \ldots, x_{n_1}$, so that the restrictions in $\boldsymbol{X}$ imply $0 \leqslant x_1 \leqslant x_2 \leqslant \cdots \leqslant x_{n_1} \leqslant 1$. Then, instead of using the $2n_1$ standard product factors x_j and $1-x_j$ for $j=1,\ldots,n_1$, we can employ the n_1+1 functions $f_k(\boldsymbol{x})$, where $f_1(\boldsymbol{x})=x_1$, $f_k(\boldsymbol{x})=x_k-x_{k-1}$ for $k=2,\ldots,n_1$, and $f_{n_1+1}(\boldsymbol{x})=1-x_{n_1}$ as specialized factors. Condition 1 is satisfied since the restrictions $0 \leqslant x_1 \leqslant x_2 \leqslant \cdots \leqslant x_{n_1} \leqslant 1$ are by assumption implied by $\boldsymbol{X}$. Condition 2a is satisfied for each $j=1,\ldots,n_1$, since for each such j, we have $x_j=\sum_{k=1}^{j} f_k(\boldsymbol{x})$ and $1-x_j=\sum_{k=j+1}^{n_1+1} f_k(\boldsymbol{x})$. As each variable x_j with $j=1,\ldots,n_1$ thus satisfies $0 \leqslant x_j \leqslant 1$, we have that condition 2b must hold true because each function $f_k(\boldsymbol{x})$ has at most one positive term, and this term is upper bounded by 1.

We provide below a small example to demonstrate the utility of exploiting variable upper bounding restrictions.

Example 4.2. Consider the following instance of Problem QP having $n=3$ binary variables $\boldsymbol{x}$ and no continuous variables $\boldsymbol{y}$ so that the functions $l(\boldsymbol{x},\boldsymbol{y})$, $g_1(\boldsymbol{x},\boldsymbol{y})$, $g_2(\boldsymbol{x},\boldsymbol{y})$, and $g_3(\boldsymbol{x},\boldsymbol{y})$ reduce to $l(\boldsymbol{x})$, $g_1(\boldsymbol{x})$, $g_2(\boldsymbol{x})$, and $g_3(\boldsymbol{x})$, respectively. (Similarly, the upcoming functions $h_1(\boldsymbol{x},\boldsymbol{y})$, $h_2(\boldsymbol{x},\boldsymbol{y})$, and $h_3(\boldsymbol{x},\boldsymbol{y})$ reduce to $h_1(\boldsymbol{x})$, $h_2(\boldsymbol{x})$, and $h_3(\boldsymbol{x})$, respectively, and the functions $\psi_k^{\boldsymbol{\pi}}(\boldsymbol{x},\boldsymbol{y})$ reduce to $\psi_k^{\boldsymbol{\pi}}(\boldsymbol{x})$ for $k=0,\ldots,5$.)

$$\begin{aligned}
\text{minimize}\quad & -5x_1+x_2+0x_3+(0x_2+0x_3)x_1+(4x_1+0x_3)x_2+(2x_1-2x_2)x_3\\
\text{subject to}\quad & \boldsymbol{x}\in\boldsymbol{X}\equiv\{\boldsymbol{x}\in\boldsymbol{S}=\{(x_1,x_2,x_3):2x_1-2x_2-2x_3\geqslant-3,\ x_1\geqslant 0,\ -x_1+x_2\geqslant 0,\\
& \quad -x_2\geqslant-1,\ x_3\geqslant 0,\ -x_3\geqslant-1\}: x_1,x_2,x_3 \text{ binary}\}
\end{aligned}$$

Thus, $l(\boldsymbol{x})=-5x_1+x_2+0x_3$, $g_1(\boldsymbol{x})=0x_2+0x_3$, $g_2(\boldsymbol{x})=4x_1+0x_3$, and $g_3(\boldsymbol{x})=2x_1-2x_2$. Problem QPRLT has three equations in (27): $w_{12}^1=w_{21}^1$, $w_{13}^1=w_{31}^1$, and $w_{23}^1=w_{32}^1$. Six restrictions are present in (28): $w_{12}^2=x_1-w_{12}^1$, $w_{13}^2=x_1-w_{13}^1$, $w_{23}^2=x_2-w_{23}^1$, $w_{21}^2=x_2-w_{21}^1$, $w_{31}^2=x_3-w_{31}^1$, and $w_{32}^2=x_3-w_{32}^1$, with no restrictions in (29) since no variables $\boldsymbol{y}$ exist. The optimal objective function value to the continuous relaxation of QPRLT is $-\frac{3}{7}$, with $(x_1,x_2,x_3)=(\frac{3}{7},\frac{6}{7},\frac{6}{7})$. An optimal dual solution for the three constraints of (27) is $\alpha_{12}^1=\frac{26}{7}$, $\alpha_{13}^1=\frac{2}{7}$, and $\alpha_{23}^1=-\frac{2}{7}$, with the nonzero optimal duals to the constraints in (28) being $\alpha_{13}^2=\frac{8}{7}$, $\alpha_{23}^2=-\frac{8}{7}$, and $\alpha_{21}^2=\alpha_{31}^2=\frac{2}{7}$. Theorem 3 ensures that the optimal objective value to Problem NP of (22) has $v^*=-\frac{3}{7}$, and that this value is realized when these dual values to (27) and (28) define $\boldsymbol{\alpha}$. The associated representation G2($\boldsymbol{\alpha}$) is as follows, where the unnecessary righthand inequalities in (14)–(17) are not listed.

$$\begin{aligned}
\text{G2}(\boldsymbol{\alpha}):\ \text{minimize}\quad & -\frac{27}{7}x_1+\frac{1}{7}x_2+\frac{2}{7}x_3+z_1^1+z_2^1+z_3^1+z_1^2+z_2^2+z_3^2\\
\text{subject to}\quad & \frac{24}{7}x_1\leqslant z_1^1\\
& \frac{24}{7}x_2+0x_3-\frac{24}{7}(1-x_1)\leqslant z_1^1\\
& -\frac{1}{7}x_2\leqslant z_2^1\\
& \frac{2}{7}x_1-\frac{2}{7}x_3-0(1-x_2)\leqslant z_2^1\\
& -\frac{2}{7}x_3\leqslant z_3^1\\
& \frac{4}{7}x_1-\frac{4}{7}x_2-0(1-x_3)\leqslant z_3^1\\
& -\frac{3}{7}(1-x_1)\leqslant z_1^2\\
& -\frac{2}{7}x_2-\frac{2}{7}x_3+\frac{2}{7}x_1\leqslant z_1^2\\
& 0(1-x_2)\leqslant z_2^2\\
& 0x_1+0x_3-0x_2\leqslant z_2^2\\
& 0(1-x_3)\leqslant z_3^2\\
& -\frac{8}{7}x_1+\frac{8}{7}x_2-\frac{4}{7}x_3\leqslant z_3^2\\
& \boldsymbol{x}\in\boldsymbol{X}
\end{aligned}$$

Here, by (11)–(13) we have $l^{\boldsymbol{\alpha}}(\boldsymbol{x}) = -\frac{27}{7}x_1 + \frac{1}{7}x_2 + \frac{2}{7}x_3$, $g_1^{\boldsymbol{\alpha}}(\boldsymbol{x}) = \frac{24}{7}x_2 + 0x_3$, $g_2^{\boldsymbol{\alpha}}(\boldsymbol{x}) = \frac{2}{7}x_1 - \frac{2}{7}x_3$, $g_3^{\boldsymbol{\alpha}}(\boldsymbol{x}) = \frac{4}{7}x_1 - \frac{4}{7}x_2$, $h_1^{\boldsymbol{\alpha}}(\boldsymbol{x}) = -\frac{2}{7}x_2 - \frac{2}{7}x_3$, $h_2^{\boldsymbol{\alpha}}(\boldsymbol{x}) = 0x_1 + 0x_3$, and $h_3^{\boldsymbol{\alpha}}(\boldsymbol{x}) = -\frac{8}{7}x_1 + \frac{8}{7}x_2$. Also, (18)–(21), with $\boldsymbol{X}^R = \boldsymbol{S}$, give $(L_1^{\boldsymbol{\alpha}1}, L_2^{\boldsymbol{\alpha}1}, L_3^{\boldsymbol{\alpha}1}) = (\frac{24}{7}, -\frac{1}{7}, -\frac{2}{7})$, $(U_1^{\boldsymbol{\alpha}0}, U_2^{\boldsymbol{\alpha}0}, U_3^{\boldsymbol{\alpha}0}) = (\frac{24}{7}, 0, 0)$, $(\bar{L}_1^{\boldsymbol{\alpha}0}, \bar{L}_2^{\boldsymbol{\alpha}0}, \bar{L}_3^{\boldsymbol{\alpha}0}) = (-\frac{3}{7}, 0, 0)$, and $(\bar{U}_1^{\boldsymbol{\alpha}1}, \bar{U}_2^{\boldsymbol{\alpha}1}, \bar{U}_3^{\boldsymbol{\alpha}1}) = (-\frac{2}{7}, 0, \frac{4}{7})$.

Now consider the special product factors $f_1(\boldsymbol{x}) = x_1$, $f_2(\boldsymbol{x}) = x_2 - x_1$, $f_3(\boldsymbol{x}) = 1 - x_2$, $f_4(\boldsymbol{x}) = x_3$, and $f_5(\boldsymbol{x}) = 1 - x_3$, which satisfy conditions 1 and 2. The optimal objective value to the continuous relaxation of SQPRLT is 0 with $(x_1, x_2) = (1, 1)$, an integer optimal. Eqs. (53) are of the form $x_1 - v_1^1 = 0$, $-x_1 + w_{12} - v_1^2 = 0$, $x_1 - w_{12} - v_1^3 = 0$, $w_{13} - v_1^4 = 0$, $x_1 - w_{13} - v_1^5 = 0$, $w_{12} - v_2^1 = 0$, $x_2 - w_{12} - v_2^2 = 0$, $-v_2^3 = 0$, $w_{23} - v_2^4 = 0$, $x_2 - w_{23} - v_2^5 = 0$, $w_{13} - v_3^1 = 0$, $-w_{13} + w_{23} - v_3^2 = 0$, $x_3 - w_{23} - v_3^3 = 0$, $x_3 - v_3^4 = 0$, and $-v_3^5 = 0$, with no equations present in (54). An optimal dual solution has the nonzero values $\pi_{12}^1 = 2$, $\pi_{22}^1 = -2$, and $\pi_{32}^1 = -2$. These computed values of $\boldsymbol{\pi}^1$give by (45) and (46) that $\psi_0^{\boldsymbol{\pi}}(\boldsymbol{x}) = -2x_1 - 2x_2$, $\psi_1^{\boldsymbol{\pi}}(\boldsymbol{x}) = \psi_3^{\boldsymbol{\pi}}(\boldsymbol{x}) = \psi_4^{\boldsymbol{\pi}}(\boldsymbol{x}) = \psi_5^{\boldsymbol{\pi}}(\boldsymbol{x}) = 0x_1 + 0x_2 + 0x_3$, and $\psi_2^{\boldsymbol{\pi}}(\boldsymbol{x}) = 2x_1 - 2x_2 - 2x_3$. To form G3($\boldsymbol{\pi}$), solve the optimization problems in (47) for k=1–5 to obtain $(\mathscr{L}_1^{\boldsymbol{\pi}}, \mathscr{L}_2^{\boldsymbol{\pi}}, \mathscr{L}_3^{\boldsymbol{\pi}}, \mathscr{L}_4^{\boldsymbol{\pi}}, \mathscr{L}_5^{\boldsymbol{\pi}}) = (0, -3, 0, 0, 0)$ and $(\mathscr{U}_1^{\boldsymbol{\pi}}, \mathscr{U}_2^{\boldsymbol{\pi}}, \mathscr{U}_3^{\boldsymbol{\pi}}, \mathscr{U}_4^{\boldsymbol{\pi}}, \mathscr{U}_5^{\boldsymbol{\pi}}) = (0, 0, 0, 0, 0)$. The following instance of G3($\boldsymbol{\pi}$) results, having an optimal objective value of 0 to the continuous relaxation so that $\eta^* = 0$ in (50) as asserted in Theorem 6.

$$
\begin{aligned}
\text{G3}(\boldsymbol{\pi}) : \text{minimize} \quad & -3x_1 + 3x_2 + z_1 + z_2 + z_3 + z_4 + z_5 \\
\text{subject to} \quad & 0x_1 \leqslant z_1 \\
& 0x_1 + 0x_2 + 0x_3 - 0(1 - x_1) \leqslant z_1 \\
& -3(x_2 - x_1) \leqslant z_2 \\
& 2x_1 - 2x_2 - 2x_3 - 0(1 + x_1 - x_2) \leqslant z_2 \\
& 0(1 - x_2) \leqslant z_3 \\
& 0x_1 + 0x_2 + 0x_3 - 0x_2 \leqslant z_3 \\
& 0x_3 \leqslant z_4 \\
& 0x_1 + 0x_2 + 0x_3 - 0(1 - x_3) \leqslant z_4 \\
& 0(1 - x_3) \leqslant z_5 \\
& 0x_1 + 0x_2 + 0x_3 - 0x_3 \leqslant z_5 \\
& \boldsymbol{x} \in \boldsymbol{X}
\end{aligned}
$$

The chosen instance of QP permits further reductions in G2($\boldsymbol{\alpha}$) and G3($\boldsymbol{\pi}$) (e.g. $z_2^2 = 0$ can be substituted from G2($\boldsymbol{\alpha}$) and $z_1 = z_3 = z_4 = z_5 = 0$ can be substituted from G3($\boldsymbol{\pi}$)). Regardless of such substitutions, transformations of variables in terms of the slacks can be used to reduce the numbers of structural constraints in both programs. In any case, G3($\boldsymbol{\pi}$) is more concise than G2($\boldsymbol{\alpha}$) and also provides a tighter relaxation.

Other functional forms $f_k(\boldsymbol{x})$ that satisfy conditions 1 and 2, and naturally arise in practice, result from generalized upper bounding restrictions. Here, the set $\boldsymbol{X}$ implies that a subset of the n variables $\boldsymbol{x}$, say $x_1, x_2, \ldots, x_{n_1}$, satisfies $\sum_{j=1}^{n_1} x_j \leqslant 1$. Then we can use the $n_1 + 1$ functions $f_k(\boldsymbol{x})$, with $f_k(\boldsymbol{x}) = x_k$ for $k = 1, \ldots, n_1$ and with $f_{n_1+1}(\boldsymbol{x}) = 1 - \sum_{j=1}^{n_1} x_j$ as specialized product factors. A similar situation arises with a special order set restriction of the form $\sum_{j=1}^{n_1} x_j = 1$, since such an equation reduces to a generalized upper bounding constraint upon treating any selected binary variable as a slack. Again, the special structure promotes a more concise formulation with a potentially tighter continuous relaxation.

5. Computational experience

Our formulations are based on a rewrite of the objective of Problem QP, together with the generation of surrogates of the constraints in QPRLT. The surrogates are motivated by ideas in [11] to maintain equivalent representations. But a question that arises is the computational performance of G2($\boldsymbol{\alpha}$) relative to the concise Problem G2 and to the larger Problem QPRLT. In particular, we are interested in the CPU times needed for Problems G2($\boldsymbol{\alpha}$) and QPRLT within a branch-and-bound framework. Although Theorem 3 tells us that G2($\boldsymbol{\alpha}$) and QPRLT have the same relaxation value when all variables are free, strength in the former can be forfeited when variables are fixed to binary values. In this section, we provide preliminary computational experience to demonstrate the potential of G2($\boldsymbol{\alpha}$) and the surrogates used in constructing this formulation.

We chose to conduct our test runs on the 0-1 quadratic knapsack problem. This problem has applications in capital budgeting, and has historically attracted research interest. It is a special case of QP where there are no continuous variables $\boldsymbol{y}$ and the set $\boldsymbol{S}$

Table 1
Computational performance

n	Problem G2			Problem QPRLT			Problem G2($\boldsymbol{\alpha}$)		
	v(G2) Gap	Nodes	CPU	v(QPRLT) Gap	Nodes	CPU	v(G2($\boldsymbol{\alpha}$)) Gap	Nodes	CPU
10	23.21	0	0	8.88	0	0	8.88	8	0
20	28.05	45	0	6.27	7	0	6.27	44	0
30	30.70	421	0	3.69	24	1	3.69	102	0
40	31.19	3899	2	3.87	185	15	3.87	826	1
50	29.65	7043	4	3.13	132	24	3.13	771	1
60	31.58	146,430	119	2.47	470	129	2.47	2559	3
70	31.71	92,967	99	2.60	662	333	2.60	4465	5
80	32.57	1,232,794	1519	2.77	877	680	2.77	8676	9
90	*	*	*	3.34	2529	2673	3.34	57,730	73
100	*	*	*	2.93	1266	2059	2.93	59,001	94

in (23) consists of a single structural inequality together with the bounding restrictions $\mathbf{0} \leqslant \boldsymbol{x} \leqslant \mathbf{1}$. It takes the form

$$\begin{aligned} \text{QKP}: \text{minimize} \quad & l(\boldsymbol{x}) + \sum_{j=1}^{n} g_j(\boldsymbol{x}) x_j \\ \text{subject to} \quad & \boldsymbol{x} \in \boldsymbol{X} \equiv \{\boldsymbol{x} \in \boldsymbol{S} : \boldsymbol{x} \text{ binary}\} \end{aligned}$$

with $l(\boldsymbol{x}) = \sum_{j=1}^{n} c_j x_j$, $g_j(\boldsymbol{x}) = \sum_{i=1, i \neq j}^{n} C_{ij} x_i \forall j = 1, \ldots, n$, and $\boldsymbol{S} \equiv \{\boldsymbol{x} : \sum_{j=1}^{n} a_j x_j \geqslant b,\ 0 \leqslant x_j \leqslant 1 \forall\ j = 1, \ldots, n\}$.

Three formulations of QKP were submitted to the mixed-integer solver of CPLEX 8.0. The first is Problem G2, where we removed the righthand inequalities of (4) and (5), and made the substitution of variables $s_j = z_j - L_j^1 x_j$ for all j in order to have only n new structural restrictions. These modifications are consistent with the observations at the end of Section 2.1, and are implemented for computational experience. The second is QPRLT, adjusted per the remarks in the closing paragraph of Section 2.2. We substituted the variables w_{ij}^1 for all $(i, j), i > j$, and w_{ij}^2 for all $(i, j), i \neq j$, out of the problem, and removed constraints (27), (28), and (30) to make this version as streamlined and competitive as possible. (The variables γ_{ij}^1 and γ_{ij}^2 and the constraints (29) are not present since there are no continuous variables in QKP.) Finally, we solved Problem G2($\boldsymbol{\alpha}$) without the righthand inequalities of (14)–(17), and upon making the substitution of variables $s_j^1 = z_j^1 - L_j^{\alpha 1} x_j$ and $s_j^2 = z_j^1 - \bar{L}_j^{\alpha 0}(1 - x_j)$ for all j. Consistent with Theorem 3, the vector $\boldsymbol{\alpha}$ was chosen as an optimal set of dual values to (27) and (28) of the relaxation of QPRLT.

The input for QKP is as follows. Motivated by [7,10,21], a_j for all j are integers taken from a uniform distribution over the interval [1, 50], and c_j for all j and C_{ij} for all (i, j) with $i < j$ are integers taken from a uniform distribution over the interval [1, 100], with C_{ij} set to C_{ji} for all $i > j$. We let $b = \frac{1}{2} \sum_{j=1}^{n} a_j$ to help ensure a consistent level of difficulty.

All tests were implemented in ANSI C++, compiled using Visual C++.Net, and executed on a Dell Workstation 340 equipped with a 2.53 GHz Pentium 4 processor and 1.5G of PC800 ECC RDRAM running Windows XP Professional. The formulations were modelled using ILOG Concert Technology 1.1.

Results are reported in Table 1 in terms of averages of ten problems, so that a total of 300 test problems are summarized. The first column records the numbers of binary variables n for 10–100 in increments of 10. The next three columns consider Problem G2, and give the gaps between the optimal binary objective *OPT* to QKP and the optimal values v(G2) to the relaxations of G2 as a percentage of *OPT*, computed as $(OPT - v(\text{G2}))/OPT \times 100$, the numbers of nodes enumerated, and the total CPU execution times in seconds. The next three columns give the same information for QPRLT for the same test problems, with the gaps between *OPT* and the optimal values v(QPRLT) to the relaxations of QPRLT computed as $(OPT - v(\text{QPRLT}))/OPT \times 100$. The final three columns repeat this same information for G2($\boldsymbol{\alpha}$). The CPU times represent all effort, including that required to compute the bounds L_j^1 and U_j^0 via the minimization problems in (6) and the maximization problems in (7), respectively, for Problem G2, and that for solving the relaxation of QPRLT (using CPLEX's Crossover Barrier Method with default settings) to obtain the desired $\boldsymbol{\alpha}$ vector as well as to compute $L_j^{\alpha 1}$, $U_j^{\alpha 0}$, $\bar{L}_j^{\alpha 0}$, and $\bar{U}_j^{\alpha 1}$ via the associated programs in (18)–(21) for G2($\boldsymbol{\alpha}$). An asterisk indicates the average solution time for the ten sample problems exceeded the 35,000 CPU second limit.

Three observations are obvious from the results of Table 1. First, Problem QPRLT has a significantly tighter relaxation value than Problem G2. Column two shows the gaps for Problem G2 ranging from 23.21% to 32.57% while columns five and eight give the gaps for QPRLT ranging from 2.47% to 8.88%. (Columns five and eight are identical by Theorem 3.) Second, Problem G2 takes less total CPU time than QPRLT for problems up to size $n = 70$, but requires more time for $n \geqslant 80$. Though Problem QPRLT examines considerably fewer nodes than G2 for all values of $n \geqslant 20$, the extra effort required to solve the tighter relaxations is not justified for the smaller-sized problems. Third, and most important to this study, Problem G2($\boldsymbol{\alpha}$) outperformed the other two formulations, never requiring more CPU time than either of these alternatives. Some strength of the relaxation values for G2($\boldsymbol{\alpha}$) was lost beyond that of QPRLT as indicated by the numbers of nodes enumerated in columns six and nine, but the effort to examine the extra nodes was more than offset by the simpler bound calculations of G2($\boldsymbol{\alpha}$), as is seen by comparing columns seven and ten.

The results of Table 1 indicate that G2($\boldsymbol{\alpha}$) is competitive with Problems G2 and QPRLT. Of course, the performance can be influenced by various factors, including problem type, input data, and strength of the relaxations of QPRLT. But the advantages to computing surrogates of constraints of QPRLT is apparent, as a means of balancing problem size and relaxation strength.

6. Conclusions

A general strategy is presented for linearizing mixed 0-1 quadratic programs so as to capture the desirable properties of concise size and tight relaxation strength within a single model. To accomplish this, two well-known linearization methods are reviewed and combined: the classical method of [11] and the level-1 representation of the reformulation-linearization technique (RLT) found in [23–25]. The first such method generates concise programs while the second promotes tight linear programming relaxations. Our study begins by enhancing the formulations in [11] using a conditional logic argument of [19,26] to adjust certain constraint coefficients, and a rewrite that alters the form of the objective function using a variable substitution based on binary identities. Both these enhancements are designed to strengthen the relaxation value.

The key observation motivating our new formulations is that the programs in [11], after applying the enhancements of conditional logic and objective rewrite, can be expressed as a type of surrogate dual of a Lagrangian subproblem of the level-1 RLT representation. The dualized constraints define the objective function rewrite, and the subproblem possesses a block-diagonal structure which inherently recognizes the strengthening due to conditional logic. Two surrogate constraints per subproblem block ensure an equivalent linearization. The objective rewrite and the surrogate constraints that combine to yield the tightest possible relaxation value are defined in terms of a computed optimal dual solution to the continuous relaxation of the level-1 RLT formulation, giving the resulting formulation the relaxation strength of the level-1 program.

Special structures within the constraints are identified that promote smaller formulations than the standard approach. One such structure arises in the general class of quadratic set partitioning problems. For this class, the level-1 RLT strength is available within a formulation of the type [11] enhanced via conditional logic, upon making simple transformations that strategically split, for each (i, j) pair with $i < j$, the objective coefficients on the product terms $x_i x_j$ and $x_j x_i$. Here, the dualized constraints define an "optimal" split. Other special structures include variable and generalized upper bounding. For these type restrictions, the special-structure RLT theory of [26] leads to more concise, tighter level-1 RLT representations than the standard RLT, which in turn motivates more concise and tighter versions of [11].

The results in this paper are of theoretical interest because they tie together two different linearization methods, and because they demonstrate how to combine the positive attributes of both methods within one formulation. But it is important to be able to use these new programs to more effectively solve nonlinear mixed 0-1 problems. We presented preliminary computational experience on the 0-1 quadratic knapsack problem to demonstrate the potential of such formulations, and believe that improved algorithms for general and specially structured nonlinear programs can be devised. As an example, formulation [11] tends not to work well on the quadratic assignment problem (QAP) due to the weak relaxation strength [16]. The level-1 RLT, however, has promoted state-of-the-art exact solution algorithms [14], even though the larger linear representations must be repeatedly solved. The linear formulation found herein for the QAP, which is a special case of the structured quadratic set partitioning problem, realizes the strength of the level-1 representation with greatly reduced size. Our ongoing research includes designing an exact algorithm for the QAP that uses these concise representations while exploiting the assignment structure in the branching process.

Acknowledgements

The first two authors would like to thank the Office of Naval Research for partially supporting this research under grant numbers N00014-97-1-0784 and N00014-00-1-0816 under their Affordability Program.

References

[1] W. Adams, R. Forrester, A simple recipe for mixed 0-1 linearizations, Operations Research Letters, to appear.
[2] W. Adams, T. Johnson, Improved linear programming-based lower bounds for the quadratic assignment problem, DIMACS Ser. Discrete Math. Theoret. Comput. Sci. 16 (1994) 43–76.
[3] W. Adams, H. Sherali, A tight linearization and an algorithm for zero–one quadratic programming problems, Management Sci. 32 (1986) 1274–1290.
[4] W. Adams, H. Sherali, Linearization strategies for a class of zero–one mixed integer programming problems, Oper. Res. 38 (1990) 217–226.
[5] W. Adams, H. Sherali, Mixed-integer bilinear programming problems, Math. Programming 59 (1993) 279–305.
[6] J. Benders, Partitioning procedures for solving mixed-variables programming problems, Numer. Math. 4 (1962) 238–252.
[7] A. Billionnet, F. Calmels, Linear programming for the 0-1 quadratic knapsack problem, European J. Oper. Res. 92 (1996) 310–325.
[8] R. Burkard, T. Bönniger, A heuristic for quadratic Boolean programs with applications to quadratic assignment problems, European J. Oper. Res. 13 (1983) 374–386.
[9] R. Burkard, U. Derigs, Assignment and matching problems: solution methods with FORTRAN-programs, Lecture Notes in Economics and Mathematical Systems, Vol. 184, Springer, Berlin, 1980, pp. 99–148.
[10] A. Caprara, D. Pisinger, P. Toth, Exact solution of the quadratic knapsack problem, INFORMS J. Comput. 11 (1999) 125–137.
[11] F. Glover, Improved linear integer programming formulations of nonlinear integer programs, Management Sci. 22 (1975) 455–460.
[12] F. Glover, An improved MIP formulation for products of discrete and continuous variables, J. Inform. Optim. Sci. 5 (1984) 469–471.
[13] P. Hahn, T. Grant, Lower bounds for the quadratic assignment problem based upon a dual formulation, Oper. Res. 46 (1998) 912–922.
[14] P. Hahn, W. Hightower, T. Johnson, M. Guignard-Spielberg, Tree elaboration strategies in branch and bound algorithms for solving the quadratic assignment problem, Yugoslav J. Oper. Res. 11 (2001) 41–60.
[15] T. Johnson, New linear programming-based solution procedures for the quadratic assignment problem, Ph.D. Dissertation, Department of Mathematical Sciences, Clemson University, 1992.
[16] L. Kaufman, F. Broeckx, An algorithm for the quadratic assignment problem using Benders' decomposition, European J. Oper. Res. 2 (1978) 207–211.
[18] J. Krarup, P. Pruzan, Computer aided-layout design, Mathematical Programming Study, Vol. 9, North-Holland Publishing Company, Amsterdam, 1978, pp. 75–94.
[19] R. Lougee-Heimer, W. Adams, A conditional logic approach for strengthening mixed 0–1 linear programs, Ann. Oper. Res., to appear.
[20] L. Lovász, S. Schrijver, Cones of matrices and set-functions and 0–1 optimization, SIAM J. Optim. 1 (1991) 166–190.
[21] P. Michelon, L. Veilleux, Lagrangean methods for the 0-1 quadratic knapsack problem, European J. Oper. Res. 92 (1996) 326–341.
[22] C. Nugent, T. Vollmann, J. Ruml, An experimental comparison of techniques for the assignment of facilities to locations, Oper. Res. 16 (1968) 150–173.
[23] H. Sherali, W. Adams, A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems, SIAM J. Discrete Math. 3 (1990) 411–430.
[24] H. Sherali, W. Adams, A hierarchy of relaxations and convex hull characterizations for mixed-integer zero-one programming problems, Discrete Appl. Math. 52 (1994) 83–106.
[25] H. Sherali, W. Adams, A Reformulation–Linearization Technique for Solving Discrete and Continuous Nonconvex Problems, Kluwer Academic Publishers, Norwell, MA, 1999.
[26] H. Sherali, W. Adams, P. Driscoll, Exploiting special structures in constructing a hierarchy of relaxations for 0–1 mixed integer problems, Oper. Res. 46 (1998) 396–405.