Preprint manuscript No.
(will be inserted by the editor)

Projective Cutting Planes for General QP with Indicator Constraints

Ulf Friedrich · **Dennis Kreber**

Received: date / Accepted: date

Abstract General quadratic optimization problems with linear constraints and additional indicator constraints on the variables are studied. Based on the well-known perspective reformulation for mixed-integer quadratic optimization problems, projective cuts are introduced as new valid inequalities for the general problem. The key idea behind the theory of these cutting planes is the projection of the continuous variables onto the space of optimal solutions dependent on the choice of indicator variables. The advantages of projective cutting planes for practical computations are illustrated with a numerical study of uncapacitated facility location problems.

Keywords MINLP · perspective reformulation · cutting planes · outer approximation

Mathematics Subject Classification 90C11 · 90C30

1 Introduction

Many optimization scenarios require *yes* or *no* decisions that activate or deactivate processes, e.g., opening a facility or selecting a certain machine. Whenever such a process is activated, it returns profit, but, at the same time, it is bound to certain restrictions or causes activation costs. To model problems of this type in a general way, we consider quadratic problems in which the variables are coupled to indicator constraints, i.e., constraints which force the continuous variables x to zero if a corresponding binary (indicator) variable z is zero:

$$\begin{aligned} \min \quad & \frac{1}{2}x^\mathsf{T}Qx - b^\mathsf{T}x + c^\mathsf{T}z \\ \text{s.t.} \quad & z_i = 0 \Rightarrow x_i = 0 \quad \forall i = 1,\dots,n \\ & z \in \mathscr{O} \\ & Ax = w \\ & Gx \geq u \\ & z \in \{0,1\}^n,\ x \in \mathbb{R}^n \end{aligned} \tag{QIND}$$

We assume that $Q = \begin{bmatrix} q_1 \cdots q_n \end{bmatrix} \in \mathbb{R}^{n\times n}$ is a symmetric, positive definite matrix, $b, c \in \mathbb{R}^n$, and $\mathscr{O}$ is a convex set. Linear equality and inequality constraints are given by matrices $A = [a_1 \cdots a_n] \in \mathbb{R}^{m\times n}$ and $G = [g_1 \cdots g_n] \in \mathbb{R}^{l\times n}$ with right-hand sides $w \in \mathbb{R}^m$ and $u \in \mathbb{R}^l$. We assume that A has full row-rank.

Problems of this structure arise in numerous applications such as portfolio optimization with short sales (Bienstock, 1996; Bertsimas and Shioda, 2009; Markowitz, 1952), best subset selection in regression (Konno and Yamamoto, 2009; Bertsimas et al., 2016; Dong et al., 2015; Miller, 1990), optimal control (Gao and Li, 2011), filter design (Wei et al., 2013), and facility location problems (Günlük and Linderoth, 2012). Solving problem (QIND) is NP-hard since the best subset selection problem is a sub-class of (QIND), which by itself is an NP-hard problem (Natarajan, 1995). Substantial research emerged in recent years improving on the capabilities to solve problems of this kind (see for example Bonami et al., 2015).

U. Friedrich
Faculty of Mathematics, Otto von Guericke University Magdeburg, Germany
E-mail: ulf.friedrich@ovgu.de

D. Kreber
Department of Mathematics, Trier University, Germany
E-mail: kreber.dennis@gmail.com

A straightforward approach for solving problems with indicator constraints is to translate a logical constraint

$$z_i = 0 \;\Rightarrow\; x_i = 0$$

into an algebraic constraint of the form

$$-M_i z_i \leq x_i \leq M_i z_i$$

where $M_i > 0$ is a sufficiently large constant such that no feasible points are cut off. While this approach is tempting, it does require a careful choice for M_i. If M_i is chosen too large, performance of branch-and-bound solvers will significantly degrade. On the other hand, if M_i is chosen too small, feasible solutions are cut off and optimality cannot be guaranteed anymore. Hence, it is mathematical folklore that so-called Big-M formulations should be avoided if possible. Furthermore, most MIP solvers nowadays allow the direct input of logical constraints as written down in (QIND). However, using this feature oftentimes does not yield smaller runtimes compared to a Big-M formulation. So it is necessary to find special structures which increase efficiency in cases of indicator constraints. Belotti et al. (2016) study the behavior of major MIP solvers when supplied with Big-M constraints. They find that the main problem lies in the relaxations being too loose and that bound tightening, in other words dynamically adjusting M_i, helps immensely with the performance.

A well-known approach to counter this problem is the perspective reformulation (Frangioni and Gentile, 2006, 2007; Günlük and Linderoth, 2010, 2012). Here, M_i is treated as a variable of the minimization problem such that M_i is forced to be as tight as possible. This method turned out to be quite effective, however it requires "more" nonlinearity as $M_i z_i$ forms a nonlinear term. Usually, this additional nonlinearity manifests as a second-order cone problem.

While the perspective reformulation can be solved significantly faster than the original formulation (QIND) most of the time, large-scale second-order cone programs can still be difficult to solve. Hence, Frangioni and Gentile (2006) and Frangioni et al. (2011) propose to add perspective cuts derived from the individual indicator constraints during the optimization of (QIND). However, this approach requires one cut per indicator constraint, which accounts for n many cuts in our setting. Furthermore, it does not consider couplings effects between the blocks.

1.1 Projective Cuts

To handle these issues, we propose cutting planes derived from projecting the continuous variables x onto the space of optimal solutions when considering z as fixed parameters, i.e., we replace the variable x with its optimal solution $\hat{x}(z)$ dependent on z. Therby, we do not face the problem of bounding the continuous variables by a value M_i. We can then derive tangents at arbitrary points z and utilize them as cutting planes for (QIND) or solve the problem by an outer approximation. As mentioned above, the perspective reformulation forces the bounds M_i to be as tight as possible. Intuition tells us that replacing x with its optimal values $\hat{x}(z)$ yields a formulation as strong as the perspective reformulation. Indeed, we show that the resulting cutting planes restrict a feasible region which coincides with the relaxation of the perspective reformulation, see Theorem 2.

Our approach has several algorithmic advantages over the perspective reformulation: First, using the perspective reformulation requires solving a second-order cone problem, and hence, introducing more computational complexity per branch-and-bound node, which our method avoids. Second, an outer approximation with tangent cuts at binary points is a "simple" MILP comprising only the indicator variables z. In addition, problems of the form (QIND) often have cardinality constraints. Hence, sparsity can be leveraged and taken advantage of.

The idea of projected tangent cuts has been studied for various special cases: Bertsimas and van Parys (2020) apply the idea to the best subset selection problem with great results. The best subset selection is a fitting candidate because it does not have any restrictions on x. Thus, it is quite straightforward to determine the projection $\hat{x}(z)$. They empirically show that an outer approximation can find solutions very fast. Bertsimas and Cory-Wright (2021) apply a similar method to the portfolio optimization problem. Here the constraints on x are more complicated and projecting x takes much more effort. For this reason the authors consider the dual of the problem restricted to the variables x. The resulting sub-problem that needs to be solved for each cut has a sophisticated formulation, but can be solved fast. Although the constraints the authors consider allow for a more general analysis, their efforts are concentrated solely on the portfolio optimization problem. Bertsimas et al. (2021) generalize the idea to nonlinear problems with indicator constraints. They as well take a duality perspective and provide a framework for nonlinear problems with indicator constraints. Unlike our approach, they provide a broad overview and hence, each indicator problem still needs special care in how the dual problem is solved, which again can be a challenging task.

Our work aims to close the gap between the general perspective that cannot provide algorithmic details and the exposition of special cases. In addition,compared to the work of Bertsimas et al. (2021) we take a different route in order to derive projected tangent cuts. We effectively work on the primal problem and handle the constraints by introducing penalization terms, which allows us to arrive at a simple cut generation procedure.

1.2 Structure of this work

In the next section, we begin our theoretical exposition by studying a simplified version of problem (QIND) that does not have the additional inequality constraints $Gx \geq u$ and equality constraints $Ax = w$. We derive a new relaxation for the problem, the matrix fractional formulation (MF). Since (MF) is not convex, we relax this problem further in Section 3 to obtain a convex relaxation (CMF). The main result of these two sections is that (CMF) is a strong as the well-known perspective reformulation.

In Section 4 we carefully study partial derivatives to derive a closed form representation for a set of valid inequalities which we call *projective cuts* for (CMF). The detailed calculations carried out in this section allow us to directly implement our findings. However, before discussing computational aspects, we turn back to the general problem (QIND) and show how to extend the results on projective cuts to problems with additional inequality and equality constraints in Section 5. The mathematical reasoning for this extension is based on a penalization approach and an analysis of the Lagrangian multipliers for growing penalty parameters. This allows us to formulate and proof our main theoretical result Theorem 6 on projective cuts for the original problem (QIND).

Finally, the computational study of Section 6 shows that our theoretical findings have a significant potential for practical computations as well. The performance analysis is based on a comparison of an outer approximation algorithm using the new projective cuts and the perspective reformulation for a set of quadratic uncapacitated facility location problems.

2 Matrix fractional reformulation

We need to reformulate the original problem for determining a projection from the variables x onto $\hat{x}(z)$. As this reformulation is rather tedious for the general case with linear equality and inequality constraints, we first focus on (QIND) without these additional constraints in the variables x. In Section 5, we then extend our findings to the general case with additional equality and inequality constraints. Therefore, we first consider the problem

$$\begin{aligned} \min \quad & \frac{1}{2}x^\mathsf{T}Qx - b^\mathsf{T}x + c^\mathsf{T}z \\ \text{s.t.} \quad & z_i = 0 \Rightarrow x_i = 0 \quad \forall i = 1,\dots,n \\ & z \in \mathscr{O} \\ & z \in \{0,1\}^n,\ x \in \mathbb{R}^n. \end{aligned} \tag{1}$$

Let $D = \operatorname{diag}(d_1,\dots,d_n)$ be a positive definite matrix that satisfies $Q - D \succcurlyeq 0$. Then, (1) can be equivalently formulated as

$$\begin{aligned} \min \quad & \frac{1}{2}x^\mathsf{T}H^\mathsf{T}Hx + \frac{1}{2}x^\mathsf{T}Dx - b^\mathsf{T}x + c^\mathsf{T}z \\ \text{s.t.} \quad & z_i = 0 \Rightarrow x_i = 0 \quad \forall i = 1,\dots,n \\ & z \in \mathscr{O} \\ & z \in \{0,1\}^n,\ x \in \mathbb{R}^n, \end{aligned} \tag{2}$$

where $H^\mathsf{T}H$ is the Cholesky decomposition of $Q - D$. We denote the columns of H by $h^1,\dots,h^n$. Isolating separable quadratic term enables the use of the perspective function (Frangioni and Gentile, 2006; Frangioni et al., 2011; Günlük and Linderoth, 2010, 2012) to strengthen a relaxation of (2). For this purpose, new variables $\tau_i := x_i^2$ are introduced and problem (2) is reformulated to the mixed-integer second-order cone program

$$\begin{aligned} \min \quad & \frac{1}{2}x^\mathsf{T}H^\mathsf{T}Hx + \frac{1}{2}\sum_{i=1}^{n} d_i\tau_i - b^\mathsf{T}x + c^\mathsf{T}z \\ \text{s.t.} \quad & x_i^2 \leq \tau_i z_i \quad \forall i = 1,\dots,n \\ & z \in \mathscr{O} \\ & z \in \{0,1\}^n,\ x \in \mathbb{R}^n. \end{aligned} \tag{PERSP}$$

The strength of the formulation relies on the choice of D. Zheng et al. (2014) present a semi-definite program to optimally choose D in order to receive the tightest relaxation of (PERSP). Dong et al. (2015) propose a similar approach for the best subset selection problem in regression. Frangioni and Gentile (2007) develop a heuristic approach, which yields a good D and requires the solution of a less computational demanding SDP.

We first reformulate the quadratic problem with indicator constraints to a matrix fractional problem akin to the work of Bertsimas and van Parys (2020), who use it in the context of the best subset selection problem. The

matrix fractional problem is highly nonlinear and not compliant with any of the prominent commercial solvers such as CPLEX or Gurobi. However, it has the advantage that tangent planes of the whole set of the resulting relaxation can be derived easily. For the aforementioned reformulation we have to assume that b lies in the span of H^T. Unfortunately, this assumption comes only with a loss of generality since $Q-D$, and therefore $H^\mathsf{T}H$, can be positive semidefinite. Hence, we alter the problem formulation slightly in this section and define $D_\varepsilon = D - \varepsilon I$ and accordingly $H_\varepsilon^\mathsf{T} H_\varepsilon = Q - D_\varepsilon$ with $H_\varepsilon =: \left[h_\varepsilon^1 \ \dots \ h_\varepsilon^n\right]$. The perturbation parameter $\varepsilon > 0$ is chosen in such a way that H_ε^T has full column rank. This perturbation yields the auxiliary problem

$$\begin{aligned} \min \quad & \frac{1}{2}x^\mathsf{T} H_\varepsilon^\mathsf{T} H_\varepsilon x + \frac{1}{2}x^\mathsf{T} D_\varepsilon x - b^\mathsf{T} x + c^\mathsf{T} z \\ \text{s.t.} \quad & z_i = 0 \Rightarrow x_i = 0 \quad \forall i = 1,\dots,n \\ & z \in \mathscr{O} \\ & z \in \{0,1\}^n,\ x \in \mathbb{R}^n. \end{aligned} \tag{3}$$

In Section 4.1 we recover the original problem (2) as ε converges to zero and derive cutting planes for (QIND).

To be able to relax problem (3) we have to represent the logical constraints as continuous terms. The usual fix for this issue is to replace the logical formulation by Big-M terms. However, as discussed in the introduction, this can lead to a weak relaxation. We approach this issue in a different way and consider z as scalars to the columns of H_ε and to the entries of b. With that intent, we denote $H_\varepsilon(z) \coloneqq H_\varepsilon \operatorname{diag}(z)$ and $b(z) \coloneqq z \circ b$, where $\circ$ denotes the Hadamard product, i.e., the component-wise multiplication. We receive the equivalent formulation

$$\begin{aligned} \min \quad & \phi_\varepsilon(x,z) \coloneqq \frac{1}{2}x^\mathsf{T} H_\varepsilon(z)^\mathsf{T} H_\varepsilon(z) x + \frac{1}{2}x^\mathsf{T} D_\varepsilon x - b(z)^\mathsf{T} x + c^\mathsf{T} z \\ \text{s.t.} \quad & z \in \mathscr{O} \\ & z \in \{0,1\}^n,\ x \in \mathbb{R}^n. \end{aligned} \tag{S}$$

Since the extracted diagonal matrix D_ε still penalizes x, a coefficient x_i shrinks as z_i goes down to 0 and vanishes at $z_i = 0$. From (S) we can naturally derive a relaxation by simply replacing $z \in \{0,1\}^n$ with $z \in [0,1]^n$. Furthermore, we denote the solution of $\min_{x\in\mathbb{R}^n} \phi_\varepsilon(x,z)$ for some fixed $z \in \mathscr{O} \cap [0,1]$ by $\hat{x}_\varepsilon(z)$. By the Karush-Kuhn-Tucker (KKT) conditions we have that

$$\hat{x}_\varepsilon(z) = \left(H_\varepsilon(z)^\mathsf{T} H_\varepsilon(z) + D_\varepsilon\right)^{-1} b(z). \tag{4}$$

By considering the KKT conditions for $\min_{x\in\mathbb{R}^n} \phi_\varepsilon(x,z)$ it is straightforward that $z_i = 0$ implies $\hat{x}_\varepsilon(z)_i = 0$. Note that $\phi_\varepsilon(x,z)$ and $\hat{x}_\varepsilon(z)$ are also well-defined for $\varepsilon = 0$. Since a perturbation of zero can be considered the most original formulation, we drop the subscript and write $\phi(x,z)$ and $\hat{x}(z)$ for the edge case $\varepsilon = 0$.

We introduced the parameter ε to guarantee that H_ε^T has full row rank. Thus, there exists a y_ε such that $b = H_\varepsilon^\mathsf{T} y_\varepsilon$ and we define the matix fractional problem

$$\begin{aligned} \min \quad & \psi_\varepsilon(z) \coloneqq \frac{1}{2} y_\varepsilon^\mathsf{T} \left(I + \sum_{i=1}^n \frac{1}{d_i - \varepsilon} h_\varepsilon^i \left(h_\varepsilon^i\right)^\mathsf{T} z_i^2 \right)^{-1} y_\varepsilon - \frac{1}{2} y_\varepsilon^\mathsf{T} y_\varepsilon + c^\mathsf{T} z \\ \text{s.t.} \quad & z \in \mathscr{O} \\ & z \in \{0,1\}^n. \end{aligned} \tag{MF}$$

It turns out that we have developed two equivalent formulations (S) and (MF) for the original problem.

Theorem 1 *Problems* (S) *and* (MF) *are equivalent and their respective relaxations are equally strong.*

Proof We show that for all $\varepsilon > 0$ the two sets

$$\{(\eta, z) \in \mathbb{R} \times [0,1]^n : \eta \geq \min_{x\in\mathbb{R}^n} \phi_\varepsilon(x,z),\ z \in \mathscr{O}\}$$

and

$$\{(\eta, z) \in \mathbb{R} \times [0,1]^n : \eta \geq \psi_\varepsilon(z),\ z \in \mathscr{O}\}$$

are equal. First, we evaluate ϕ_ε at $\hat{x}_\varepsilon(z)$ to receive the equivalent matrix fractional formulation. Accordingly, it holds that

$$\phi_\varepsilon(\hat{x}_\varepsilon(z), z) = -\frac{1}{2} b(z)^\mathsf{T} \left(H_\varepsilon(z)^\mathsf{T} H_\varepsilon(z) + D_\varepsilon\right)^{-1} b(z) + c^\mathsf{T} z.$$

Since $b = H_\varepsilon^\mathsf{T} y_\varepsilon$, and hence $b(z) = H_\varepsilon(z)^\mathsf{T} y_\varepsilon$, we can apply the Sherman-Morrison-Woodbury identity (Meyer, 2000, p. 124)

$$\begin{aligned}
&-\frac{1}{2} b(z)^\mathsf{T} \left(H_\varepsilon(z)^\mathsf{T} H_\varepsilon(z) + D_\varepsilon \right)^{-1} b(z) + c^\mathsf{T} z \\
&= -\frac{1}{2} y_\varepsilon^\mathsf{T} H_\varepsilon(z) \left(H_\varepsilon(z)^\mathsf{T} H_\varepsilon(z) + D_\varepsilon \right)^{-1} H_\varepsilon(z)^\mathsf{T} y_\varepsilon + c^\mathsf{T} z \\
&= \frac{1}{2} y_\varepsilon^\mathsf{T} \left(I + H_\varepsilon(z) D_\varepsilon^{-1} H_\varepsilon(z)^\mathsf{T} \right) y_\varepsilon - \frac{1}{2} y_\varepsilon^\mathsf{T} y_\varepsilon + c^\mathsf{T} z.
\end{aligned}$$

Decomposing $H_\varepsilon(z) D_\varepsilon^{-1} H_\varepsilon(z)^\mathsf{T}$ into a sum of rank-1 matrices yields

$$H_\varepsilon(z) D_\varepsilon^{-1} H_\varepsilon(z)^\mathsf{T} = \sum_{i=1}^{n} \frac{1}{d_i - \varepsilon} h_\varepsilon^i \left(h_\varepsilon^i \right)^\mathsf{T} z_i^2$$

and thus $\min_{x \in \mathbb{R}^n} \phi_\varepsilon(x,z) = \psi_\varepsilon(z)$ follows, which proves the theorem. □

The function $\psi_\varepsilon(z)$ is non-convex in general since for $n = 1$ we can choose the parameters such that $\psi_\varepsilon(z) = \frac{1}{1+z^2}$, which can easily been verified to be non-convex. We obtain a convex formulation by replacing z_i^2 with z_i. This transformation has no effect on binary solutions and hence the following *convexified matrix fractional formulation*

$$\begin{aligned}
\min \quad & \zeta_\varepsilon(z) := \frac{1}{2} y_\varepsilon^\mathsf{T} \left(I + \sum_{i=1}^{n} \frac{1}{d_i - \varepsilon} h_\varepsilon^i \left(h_\varepsilon^i \right)^\mathsf{T} z_i \right)^{-1} y_\varepsilon - \frac{1}{2} y_\varepsilon^\mathsf{T} y_\varepsilon + c^\mathsf{T} z \\
\text{s.t.} \quad & z \in \mathscr{O} \\
& z \in \{0,1\}^n
\end{aligned} \tag{CMF}$$

is equivalent to (MF) and (S). While the relaxation of this formulation is of course weaker than the relaxations of the two formulations of Theorem 1, it is convex (Burke and Hoheisel, 2015).

Since ζ_ε is highly non-linear and not compatible with the prominent mixed-integer black box solvers, we intend to derive tangent planes of the function ζ_ε and utilize them as cutting planes. However, to see that our intention is sensible, i.e., that the cuts we gain are actually effective, we show that (CMF) is equivalent to the perspective reformulation akin to (PERSP). The perspective reformulation is known to be very effective (Frangioni and Gentile, 2006; Dong et al., 2015; Atamtürk and Gómez, 2018) and the cuts should provide the same effectivity while not requiring a second-order cone formulation, which is infamous for becoming hard to solve for large-scale instances.

3 Equivalence of the perspective reformulation and (CMF)

The tight relationship between the perspective reformulation (PERSP) and the matrix fractional problem (CMF) has been observed previously, but there seems to be a lack of awareness for this connection. Bertsimas and van Parys (2020) applied projected tangent planes to the best subset selection problem but seem to be unaware of the similarity. To the best of our knowledge the first explanations of the equivalence are made by Kreber (2019) and Xie and Deng (2020). The topic is also discussed by Bertsimas and Cory-Wright (2021) and Bertsimas et al. (2021). We want to revisit the connection between the problems and show that they are indeed similar.

For $z \in \mathbb{R}^n_+$ let us consider the function

$$M(z) = M_0 + \sum_{i=1}^{n} M_i z_i$$

with $M_i \in \mathbb{R}^{n \times n}$, $i = 0, \ldots, n$, being positive semi-definite. Therefore, $M(z)$ is also positive semi-definite. Then, the epigraph of

$$\begin{aligned}
f : \mathbb{R}^n \times \mathbb{R}^n_+ &\to \mathbb{R} \\
f(y,z) &\mapsto y^\mathsf{T} \left(M(z) \right)^{-1} y
\end{aligned}$$

is second-order cone representable (Nesterov and Nemirovskii, 1994, pp. 227 - 229), i.e., there exists a second-order cone formulation whose feasible region coincidences with the set $R \coloneqq \{(\eta, y, z) \in \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^p_+ : \eta \geq f(y,z)\}$. More precisely, the feasible region of

$$\begin{aligned}
\sum_{i=0}^{n} M_i^{\frac{1}{2}} u_i &= y \\
\|u_i\|_2^2 &\leq z_i \tau_i \qquad \forall i \in \{0,\dots,n\} \\
\sum_{i=0}^{n} \tau_i &\leq \eta \\
z_0 &= 1 \\
u_i \in \mathbb{R}^n, \tau_i \in \mathbb{R}_+ &\qquad \forall i \in \{0,\dots,n\} \\
\eta \in \mathbb{R},\ x \in \mathbb{R}^n,\ z \in \mathbb{R}^n_+ &
\end{aligned}$$

is equal to R. Using this identity with (CMF) yields

$$\begin{aligned}
u_0 + \sum_{i=1}^{n} \left(\frac{1}{d_i - \varepsilon}\right)^{\frac{1}{2}} \|h_\varepsilon^i\|_2^{-1} h_\varepsilon^i \left(h_\varepsilon^i\right)^{\mathsf{T}} u_i &= \frac{1}{\sqrt{2}} y_\varepsilon \\
\|u_i\|_2^2 &\leq z_i \tau_i \qquad \forall i \in \{1,\dots,n\} \\
\|u_0\|_2^2 &\leq \tau_0 \\
\sum_{i=0}^{n} \tau_i - \frac{1}{2} y_\varepsilon^{\mathsf{T}} y_\varepsilon + c^{\mathsf{T}} z &\leq \eta \\
z &\in \mathscr{O} \\
u_i \in \mathbb{R}^n, \tau_i \in \mathbb{R}_+ &\qquad \forall i \in \{0,\dots,n\} \\
\eta \in \mathbb{R},\ x \in \mathbb{R}^n,\ z \in [0,1]^n. &
\end{aligned} \tag{5}$$

For this feasibility problem it can be shown that the formulation can be simplified further under the assumption that η is minimal. The following Lemma states this simplification. The proof is given in the appendix.

Lemma 1 *Let $\bar{z} \in \mathscr{O} \cap [0,1]^n$ be fixed and let $(\hat{\eta}, \hat{\tau}, \hat{u}_0, \dots, \hat{u}_n)$ be a solution of*

$$\begin{aligned}
&\min && \eta \\
&\text{s.t.} && (\eta, \bar{z}, \tau, u_0, \dots, u_n) \text{ is a feasible point of (5)}
\end{aligned} \tag{6}$$

then there exist $\mu_1, \dots \mu_n \in \mathbb{R}$ such that $\hat{u}_i = \mu_i h_\varepsilon^i$ for every $i = 1,\dots,n$.

Under the assumption that we seek to minimize η, we can replace u_i with $\mu_i h_\varepsilon^i$ for every $i = 1,\dots,n$. Substituting $x_i \coloneqq \left(\frac{2}{d_i - \varepsilon}\right)^{\frac{1}{2}} \|h_\varepsilon^i\|_2 \mu_i$ and canceling out u_0 leads to the program

$$\begin{aligned}
\frac{1}{2} x^{\mathsf{T}} H_\varepsilon^{\mathsf{T}} H_\varepsilon x + \sum_{i=1}^{n} \tau_i - b^{\mathsf{T}} x + c^{\mathsf{T}} x &\leq \eta \\
\frac{d_i - \varepsilon}{2} x_i^2 &\leq z_i \tau_i \qquad \forall i \in \{1,\dots,n\} \\
z &\in \mathscr{O} \\
\eta \in \mathbb{R},\ x \in \mathbb{R}^n,\ \tau \in \mathbb{R}^n_+,\ z \in [0,1]^n &
\end{aligned}$$

which is the perspective reformulation (PERSP). Clearly, the following theorem follows immediately from the observations in this section.

Theorem 2 *The convexified matrix fractional formulation* (CMF) *is equivalent to the perspective reformulation* (PERSP).

The perspective reformulation is known to be a strong relaxation. Since the (CMF) is equivalent to it by the above theorem, it is reasonable to derive cuts from (CMF).

4 Tangent cuts

Next, we derive tangent cuts from problem (CMF). Since they are valid for a relaxation of (QIND), they are also valid for the original problem. Since they originate from projecting x onto the space of optimal solutions, we are calling them *projective cuts*. The advantage of this approach is that they are as strong as the perspective reformulation while having a much lower dimension, i.e., the problem (CMF) only consists of the variables z which are exclusively binary variables. This basic version of projective cuts will be tightened in the next sections when incorporating the constraints that have been ignored in (CMF).

For deriving tangent cuts we require the gradient of ζ_ε. We first consider the following result.

Proposition 1 (Bertsimas and van Parys 2020) *Let $M : [0,1]^n \to \mathbb{R}^{n\times n}$ be a matrix function and $y \in \mathbb{R}^n$. Assume that $(I+M(z))^{-1}$ exists for all $z \in [0,1]^n$, then the gradient of $y^\mathsf{T}(I+M(z))^{-1}y$ is given by*

$$\frac{\partial}{\partial z_i} y^\mathsf{T}(I+M(z))^{-1}y = -y^\mathsf{T}(I+M(z))^{-1}\left(\frac{\partial M}{\partial z_i}\right)(I+M(z))^{-1}y.$$

We apply the proposition to the relaxation of (CMF) to obtain an explicit representation of the gradient, which we require to determine the tangents at given vectors z.

Theorem 3 *For $\varepsilon > 0$ the partial derivative of the objective function of* (CMF) *is given by*

$$\frac{\partial}{\partial z_i}\zeta_\varepsilon(z) = -\frac{\left(b_i - \left(h_\varepsilon^i\right)^\mathsf{T} H_\varepsilon(\sqrt{z})\hat{x}_\varepsilon(\sqrt{z})\right)^2}{2(d_i-\varepsilon)} + c_i$$

for all $i = 1,\ldots,n$. Moreover, for all indices i with $z_i > 0$ the identity

$$-\frac{\left(b_i - \left(h_\varepsilon^i\right)^\mathsf{T} H_\varepsilon(\sqrt{z})\hat{x}_\varepsilon(\sqrt{z})\right)^2}{2(d_i-\varepsilon)} = -\frac{d_i-\varepsilon}{2z_i}\left(\hat{x}_\varepsilon(\sqrt{z})_i\right)^2$$

holds.

Proof By Proposition 1 we have

$$\begin{aligned}
&\frac{\partial}{\partial z_i}\frac{1}{2}y_\varepsilon^\mathsf{T}\left(I+\sum_{i=1}^n \frac{1}{d_i-\varepsilon}h_\varepsilon^i\left(h_\varepsilon^i\right)^\mathsf{T} z_i\right)^{-1} y_\varepsilon \\
&= -\frac{1}{2}y_\varepsilon^\mathsf{T}\left(I+\sum_{i=1}^n \frac{1}{d_i-\varepsilon}h_\varepsilon^i\left(h_\varepsilon^i\right)^\mathsf{T} z_i\right)^{-1}\left(\frac{1}{d_i-\varepsilon}h_\varepsilon^i\left(h_\varepsilon^i\right)^\mathsf{T}\right) \\
&\quad\cdot\left(I+\sum_{i=1}^n \frac{1}{d_i-\varepsilon}h_\varepsilon^i\left(h_\varepsilon^i\right)^\mathsf{T} z_i\right)^{-1} y_\varepsilon \\
&= -\frac{1}{2(d_i-\varepsilon)}\left(\left(h_\varepsilon^i\right)^\mathsf{T}\left(I+\sum_{i=1}^n \frac{1}{d_i-\varepsilon}\sqrt{z_i}h_\varepsilon^i\left(h_\varepsilon^i\right)^\mathsf{T}\sqrt{z_i}\right)^{-1} y_\varepsilon\right)^2.
\end{aligned}$$

Applying the Sherman-Morrison-Woodbury matrix identity yields

$$\begin{aligned}
&-\frac{1}{2(d_i-\varepsilon)}\left(\left(h_\varepsilon^i\right)_i^\mathsf{T} y_\varepsilon - \left(h_\varepsilon^i\right)^\mathsf{T} H_\varepsilon(\sqrt{z})\left(D_\varepsilon + H_\varepsilon(\sqrt{z})^\mathsf{T} H_\varepsilon(\sqrt{s})\right)^{-1} H_\varepsilon(\sqrt{z})^\mathsf{T} y_\varepsilon\right)^2 \\
=&-\frac{1}{2(d_i-\varepsilon)}\left(b_i - \left(h_\varepsilon^i\right)^\mathsf{T} H_\varepsilon(\sqrt{z})\left(D_\varepsilon + H_\varepsilon(\sqrt{z})^\mathsf{T} H_\varepsilon(\sqrt{z})\right)^{-1} b(\sqrt{z})\right)^2.
\end{aligned}$$

By the definition of $\hat{x}_\varepsilon$ it follows that

$$\begin{aligned}
&-\frac{1}{2(d_i-\varepsilon)}\left(b_i - \left(h_\varepsilon^i\right)^\mathsf{T} H_\varepsilon(\sqrt{z})\left(D_\varepsilon + H_\varepsilon(\sqrt{z})^\mathsf{T} H_\varepsilon(\sqrt{z})\right)^{-1} b(\sqrt{z})\right)^2 \\
=&-\frac{1}{2(d_i-\varepsilon)}\left(b_i - \left(h_\varepsilon^i\right)^\mathsf{T} H_\varepsilon(\sqrt{z})\hat{x}_\varepsilon(\sqrt{z})\right)^2,
\end{aligned}$$

which proves the first statement of the theorem. Further, it can easily be seen that for $z_i > 0$ it holds that

$$\begin{aligned}
&-\frac{1}{2(d_i-\varepsilon)}\left(b_i-\left(h_\varepsilon^i\right)^\mathsf{T} H_\varepsilon(\sqrt{z})\hat{x}_\varepsilon(\sqrt{z})\right)^2\\
=&-\frac{1}{2(d_i-\varepsilon)}\left(b_i-\frac{1}{\sqrt{z_i}}\left(\sqrt{z_i}\left(h_\varepsilon^i\right)^\mathsf{T} H_\varepsilon(\sqrt{z})+(d_i-\varepsilon)e_i^\mathsf{T}\right)\hat{x}_\varepsilon(\sqrt{z})+\frac{d_i-\varepsilon}{\sqrt{z_i}}\hat{x}_\varepsilon(\sqrt{z})_i\right)^2\\
=&-\frac{1}{2(d_i-\varepsilon)}\left(b_i-\frac{1}{\sqrt{z_i}}\left(b(\sqrt{z})\right)_i+\frac{d_i-\varepsilon}{\sqrt{z_i}}\hat{x}_\varepsilon(\sqrt{z})_i\right)^2\\
=&-\frac{d_i-\varepsilon}{2z_i}\left(\hat{x}_\varepsilon(\sqrt{z})_i\right)^2,
\end{aligned}$$

where e_i is the i-th unit vector. □

We now have all the technical results needed to derive valid tangent cuts for problem (S). Let us recapitulate the analytic steps we have taken so far: We first introduced (S), which is a natural representation of the underlying idea. That is, the variables x_i are scaled by z_i and simultaneously penalized by the diagonal matrix D such that they are forced to decrease when z_i decreases. We then projected the continuous variables onto the set of minima with respect to the parameters z and received the matrix fractional problem (MF). Unfortunately, (MF) is not convex, which is why we relaxed it to problem (CMF). For this problem we determined the gradient above. We arrive at the main result of this section:

Theorem 4 *Let be $\bar{z} \in \mathscr{O} \cap [0,1]^n$ and $\varepsilon > 0$. The inequality*

$$\eta \geq \zeta_\varepsilon(\bar{z}) + \nabla\zeta_\varepsilon(\bar{z})\,(z-\bar{z}) \tag{7}$$

is valid for the set of feasible points in

$$\begin{aligned}
&\eta \geq \frac{1}{2}x^\mathsf{T}Qx - b^\mathsf{T}x + c^\mathsf{T}z\\
&z_i = 0 \Rightarrow x_i = 0 \qquad\qquad \forall i = 1,\dots,n\\
&z \in \mathscr{O}\\
&z \in \{0,1\}^n,\ x \in \mathbb{R}^n, \eta \in \mathbb{R}_+.
\end{aligned}$$

Even though (7) is valid for the original problem (QIND) it would be preferable to find tangent cuts without the parameter ε. After all, D was supposed to be chosen such that the relaxation is the tightest. Hence, we want do avoid perturbing the matrix by ε. Therefore, we next consider the limit of (7) when ε converges to 0.

4.1 Independence of the parameter ε

In this section we revisit the cutting plane (7) for $\varepsilon > 0$ becoming arbitrarily small. We consider every component of (7), i.e., $\hat{x}_\varepsilon, \zeta_\varepsilon$, and $\nabla\zeta_\varepsilon$, and determine the limit for each of them. Thus, we can derive a valid cutting plane of the form (7) also for $\varepsilon = 0$ at the end of the section. We first examine the limit of $\hat{x}_\varepsilon$.

Lemma 2 *Let z be an element of $\mathscr{O} \cap [0,1]^n$. Then, the following limits hold*

i) $\lim\limits_{\varepsilon\to 0} \hat{x}_\varepsilon(z) = \hat{x}(z)$,

ii) $\lim\limits_{\varepsilon\to 0} \frac{\partial}{\partial z_i}\zeta_\varepsilon(z) = -\frac{\left(b_i-(q_i-d_ie_i)^\mathsf{T}(\sqrt{z}\circ\hat{x}(\sqrt{z}))\right)^2}{2d_i} + c_i$,

iii) $\lim\limits_{\varepsilon\to 0} \zeta_\varepsilon(z) = \phi(\hat{x}(\sqrt{z}), \sqrt{z}) + c^\mathsf{T}(z-\sqrt{z})$.

Moreover, the limit of the gradient can be further simplified to

$$\lim_{\varepsilon\to 0}\frac{\partial}{\partial z_i}\zeta_\varepsilon(z) = \begin{cases} -\frac{d_i}{2z_i}\left(\hat{x}(\sqrt{z})_i\right)^2 + c_i, & \text{if } z_i > 0,\\ -\frac{\left(b_i-q_i^\mathsf{T}(\sqrt{z}\circ\hat{x}(\sqrt{z}))\right)^2}{2d_i} + c_i, & \text{if } z_i = 0. \end{cases}$$

The proof of the lemma is given in Appendix B.

Finally, we study the convergence of the cutting planes (7) when ε converges to 0. We show that the resulting hyperplane is a valid cut for the original problem (QIND). For that, we define the two index sets

$$I(z) \coloneqq \{i \in \{1,\dots,n\} : z_i > 0\},$$
$$J(z) \coloneqq \{j \in \{1,\dots,n\} : z_j = 0\},$$

and the values

$$\delta_i(z) \coloneqq \frac{\left(\hat{x}(\sqrt{z})_i\right)^2}{z_i},$$
$$\gamma_i(z) \coloneqq \left(b_i - q_i^\mathsf{T}\left(\sqrt{z}\circ\hat{x}(\sqrt{z})\right)\right)^2.$$

Theorem 5 *Let be $\bar{z} \in \mathscr{O} \cap [0,1]^n$. The inequality*

$$\eta \geq \phi\left(\hat{x}(\sqrt{\bar{z}}), \sqrt{\bar{z}}\right) - \frac{1}{2}\sum_{i\in I(\bar{z})}^{n} d_i\delta_i(\bar{z})(z_i - \bar{z}_i) - \frac{1}{2}\sum_{j\in J(\bar{z})}^{n}\frac{\gamma_j(\bar{z})}{d_j}z_j + c^\mathsf{T}\left(z + \sqrt{\bar{z}}\right)$$

is valid for the set of feasible points in

$$\begin{aligned} \eta &\geq \frac{1}{2}x^\mathsf{T}Qx - b^\mathsf{T}x + c^\mathsf{T}z \\ z_i = 0 &\Rightarrow x_i = 0 \qquad\qquad \forall i = 1,\dots,n \\ z &\in \mathscr{O} \\ z &\in \{0,1\}^n,\ x \in \mathbb{R}^n,\ \eta \in \mathbb{R}. \end{aligned} \tag{8}$$

Proof Let $(\bar{\eta}, \bar{x}, \bar{z})$ be a feasible point of (8). Then, by Theorem 4, we have that for any $\varepsilon > 0$ sufficiently small the inequality

$$\eta \geq \zeta_\varepsilon(\bar{z}) + \nabla\zeta_\varepsilon(\bar{z})\,(z - \bar{z})$$

is valid. As Lemma 2 states that $\zeta_\varepsilon(\bar{z})$ and $\nabla\zeta_\varepsilon(\bar{z})$ converge with $\varepsilon \to 0$, the inequality

$$\eta \geq \lim_{\varepsilon\to 0}\zeta_\varepsilon(\bar{z}) + \nabla\zeta_\varepsilon(\bar{z})\,(z - \bar{z})$$

is valid as well. Taking the limit of both terms dependent on ε yields the statement of this theorem. □

5 Additional equality and inequality constraints

In this section, we extend our analysis to the general problem (QIND) that has constraints $Ax = w$ and $Gx \geq u$ on the continuous variables. The issue with these additional inequalities and equalities is that an explicit description of the objective value dependent on the variables z is required. To handle this difficulty, we propose to consider the penalized problem

$$\begin{aligned} \min \quad & \frac{1}{2}x^\mathsf{T}Qx - b^\mathsf{T}x + c^\mathsf{T}z + \frac{\pi}{2}\|Ax - w\|_2^2 + \frac{\pi}{2}\sum_{i=1}^{l}\max(0, u_i - (Gx)_i)^2 \\ \text{s.t.} \quad & z_i = 0 \Rightarrow x_i = 0 \quad \forall i = 1,\dots,n, \\ & z \in \mathscr{O}, \\ & z \in \{0,1\}^n,\ x \in \mathbb{R}^n, \end{aligned} \tag{PQIND}$$

which does yield a point arbitrarily close to the solution of (QIND) for sufficiently large π, see Nocedal and Wright (2006). Nevertheless, we do not intend to solve the penalized problem. Instead it serves as a theoretical aid to derive valid inequalities as it tailors the problem to fit into the framework developed in the previous part of the article. We will see that, after some manipulation, the results are indeed independent of the penalization parameter π. Let us first consider fixed x and z. We denote the set of violated constraints by

$$V(x,z) \coloneqq \left\{j \in \{1,\dots,l\} : \left(G(z\circ x)\right)_j < u_i\right\},$$

the submatrix of G with rows only in $V(x,z)$ by

$$G(x,z) \coloneqq [g_1(x,z)\cdots g_n(x,z)] \coloneqq [G_{j,\bullet}]_{j\in V(x,z)},$$

and the subvector of u with rows only in $V(x,z)$ by

$$u(x,z) \coloneqq (u_j)_{j \in V(x,z)}.$$

With this notation we rewrite formulation (PQIND) as

$$\begin{aligned} \min \quad & \frac{1}{2} x^\mathsf{T} \Big(Q + \pi A^\mathsf{T} A + \pi G(x,z)^\mathsf{T} G(x,z) - D \Big) x + x^\mathsf{T} D x \\ & - \Big(b + \pi A^\mathsf{T} w + \pi G(x,z)^\mathsf{T} u(x,z) \Big)^\mathsf{T} x + c^\mathsf{T} z \\ & + \|w\|_2^2 + \|u(x,z)\|_2^2 \\ \text{s.t.} \quad & z_i = 0 \Rightarrow x_i = 0 \quad \forall i = 1, \dots, n, \\ & z \in \mathscr{O}, \\ & z \in \{0,1\}^n,\ x \in \mathbb{R}^n, \end{aligned} \tag{9}$$

where D is again a diagonal matrix with positive entries on the diagonal such that

$$Q + \pi A^\mathsf{T} A + \pi G(x,z)^\mathsf{T} G(x,z) - D \succeq 0.$$

Since $Q + \pi A^\mathsf{T} A + \pi G(x,z)^\mathsf{T} G(x,z)$ relies on π, x, and z, it seems natural that D must also rely on π, x, and z. This, however, would complicate our intentions immensely. Fortunately, we can simply choose D such that $Q - D \succeq 0$. The reason for this is that $\pi A^\mathsf{T} A$ and $\pi G(x,z)^\mathsf{T} G(x,z)$ are positive semi-definite. Hence, since the set of positive semi-definite matrices forms a cone we can simply add them to $Q - D$ without breaking the positive semi-definiteness. For a relaxation we analogously define

$$\begin{aligned} \phi_\pi^{\text{pen}}(x,z) \coloneqq & \frac{1}{2} (z \circ x)^\mathsf{T} \Big(Q + \pi A^\mathsf{T} A + \pi G(x,z)^\mathsf{T} G(x,z) - D \Big) (z \circ x) + \frac{1}{2} x^\mathsf{T} D x \\ & - \Big(b + \pi A^\mathsf{T} w + \pi G(x,z)^\mathsf{T} u(x,z) \Big)^\mathsf{T} (z \circ x) + c^\mathsf{T} z + \frac{\pi}{2} \|w\|_2^2 + \frac{\pi}{2} \|u(x,z)\|_2^2 \end{aligned}$$

and arrive at a relaxation of (PQIND)

$$\min_{\substack{x \in \mathbb{R}^n \\ z \in \mathscr{O} \cap [0,1]^n}} \quad \phi_\pi^{\text{pen}}(x,z). \tag{10}$$

We define $\hat{x}_\pi^{\text{pen}}(z)$ to be a solution to the problem $\min_{x \in \mathbb{R}^n} \phi_\pi^{\text{pen}}(x,z)$ for fixed $z \in \mathscr{O} \cap [0,1]^n$. In order to derive valid cuts, we consider the feasibility problem

$$\begin{aligned} \eta & \geq \frac{1}{2} x^\mathsf{T} Q x - b^\mathsf{T} x + c^\mathsf{T} z \\ z_i = 0 & \Rightarrow x_i = 0 \qquad \forall i = 1, \dots, n \\ z & \in \mathscr{O} \\ Ax & = w \\ Gx & \geq u \\ z & \in \{0,1\}^n,\ x \in \mathbb{R}^n,\ \eta \in \mathbb{R}, \end{aligned} \tag{QIND$_{\text{feas}}$}$$

and denote the points satisfying (QIND$_{\text{feas}}$) by F. Similarly to Section 2, we define

$$\begin{aligned} \delta_{\pi,i}^{\text{pen}}(z) \coloneqq & \frac{\big(\hat{x}_\pi^{\text{pen}}(\sqrt{z})_i\big)^2}{z_i}, \\ \gamma_{\pi,i}^{\text{pen}}(z) \coloneqq & \Big(b_i + \pi a_i^\mathsf{T} w + \pi g(\hat{x}_\pi^{\text{pen}}(\sqrt{z}), \sqrt{z})_i^\mathsf{T} u(\hat{x}(\sqrt{z}), \sqrt{z}) \\ & - \Big(q_i + \pi A^\mathsf{T} a_i + \pi G(\hat{x}_\pi^{\text{pen}}(\sqrt{z}), \sqrt{z})^\mathsf{T} g(\hat{x}_\pi^{\text{pen}}(\sqrt{z}), \sqrt{z})_i \Big)^\mathsf{T} \big(\sqrt{z} \circ \hat{x}_\pi^{\text{pen}}(\sqrt{z}) \big) \Big)^2 \end{aligned}$$

and prove the following lemma.

Lemma 3 *Let be $\bar{z} \in \mathscr{O} \cap [0,1]^n$. Then, the inequality*

$$\eta \geq \phi_\pi^{pen}\big(\hat{x}_\pi^{pen}(\sqrt{\bar{z}}), \sqrt{\bar{z}}\big) - \frac{1}{2} \sum_{i \in I(\bar{z})}^{n} d_i \delta_{\pi,i}^{pen}(\bar{z})(z_i - \bar{z}_i) - \frac{1}{2} \sum_{j \in J(\bar{z})}^{n} \frac{\gamma_{\pi,j}^{pen}(\bar{z})}{d_j} z_j + c^\mathsf{T}(z + \sqrt{\bar{z}}) \tag{11}$$

is valid for F.

Proof Assume $\bar{z}$ is chosen such that the set

$$\left\{ j \in \{1,\dots,l\} : G\left(\sqrt{\bar{z}} \circ \hat{x}_\pi^{\text{pen}}(\sqrt{\bar{z}})\right) = u \right\}$$

is empty. Then, for all z with $\|z-\bar{z}\|_2$ sufficiently small, the set $V(\hat{x}_\pi^{\text{pen}}(z),z)$ does not change. Therefore restricting z to an appropriately small open set U with $\bar{z} \in U$, we can assume that problem (10) is a quadratic program with indicator constraints and no constraints concerning x. Thus, we can apply Theorem 5 and have that (11) is a valid cut for

$$\begin{aligned} &\eta \geq \frac{1}{2}x^\mathsf{T}Qx - b^\mathsf{T}x + c^\mathsf{T}z + \frac{\pi}{2}\|Ax-w\|_2^2 + \frac{\pi}{2}\sum_{i=1}^{l}\max(0,u_i-(Gx)_i)^2 \\ &z_i = 0 \Rightarrow x_i = 0 \qquad \forall i = 1,\dots,n, \\ &z \in \mathscr{O}, \\ &z \in \{0,1\}^n,\ x \in \mathbb{R}^n. \end{aligned} \tag{12}$$

That means that inequality (11) is satisfied for every feasible point in (12). Hence, it is satisfied for the particular points (η,z,x) for which $Ax = w$ and $Gx \geq u$ hold true.

Now let us consider the case where

$$\left\{ j \in \{1,\dots,l\} : G\left(\sqrt{\bar{z}} \circ \hat{x}_\pi^{\text{pen}}(\sqrt{\bar{z}})\right) = u \right\}$$

is nonempty. That means there are inequalities which can be violated after perturbing $\bar{z}$. Hence, there exist a sufficiently small and open set U with $\bar{z} \in U$ such that $V(\hat{x}_\pi^{\text{pen}}(z),z) \supseteq V(\hat{x}_\pi^{\text{pen}}(\bar{z}),\bar{z})$ for all $z \in U$. We will now argue that we can limit the objective function to the rows indexed by $V(\hat{x}_\pi^{\text{pen}}(\bar{z}),\bar{z})$ and that a valid inequality for the reduced row indices is also a valid bound for the original objective function. Denote

$$\begin{aligned} \widetilde{G} &:= G(\hat{x}_\pi^{\text{pen}}(\sqrt{\bar{z}}),\sqrt{\bar{z}}), \\ \tilde{u} &:= u(\hat{x}_\pi^{\text{pen}}(\sqrt{\bar{z}}),\sqrt{\bar{z}}), \\ \Phi(x,z) &:= \frac{1}{2}(z\circ x)^\mathsf{T}\left(Q + \pi A^\mathsf{T}A + \pi\widetilde{G}^\mathsf{T}\widetilde{G} - D\right)(z\circ x) + \frac{1}{2}x^\mathsf{T}Dx \\ &\quad - \left(b + \pi A^\mathsf{T}w + \pi\widetilde{G}^\mathsf{T}\tilde{u}\right)^\mathsf{T}(z\circ x) + c^\mathsf{T}z + \frac{\pi}{2}\|w\|_2^2 + \frac{\pi}{2}\|\tilde{u}\|_2^2, \end{aligned}$$

i.e., we fix the set $V(\hat{x}_\pi^{\text{pen}}(\bar{z}),\bar{z})$. Now, since $V(\hat{x}_\pi^{\text{pen}}(z),z) \supseteq V(\hat{x}_\pi^{\text{pen}}(\bar{z}),\bar{z})$ holds for all $z \in U$ we have that

$$\phi_\pi^{\text{pen}}(\hat{x}_\pi^{\text{pen}}(\sqrt{z}),\sqrt{z}) \geq \Phi(\hat{x}_\pi^{\text{pen}}(\sqrt{z}),\sqrt{z})$$

for all $z \in U$. In other words, there might be more violated inequalities at the point z compared to the point $\bar{z}$, and hence, $\phi_\pi^{\text{pen}}(\hat{x}_\pi^{\text{pen}}(\sqrt{z}),\sqrt{z})$ consists of more penalty terms while the number of penalty terms is fixed for $\Phi\left(\hat{x}_\pi^{\text{pen}}(\sqrt{z}),\sqrt{z}\right)$.

It follows that an inequality bounding the epigraph of Φ will also be a lower bound for the epigraph of ϕ_π^{pen}. Hence, applying Theorem 5 to problem

$$\begin{aligned} &\eta \geq \frac{1}{2}x^\mathsf{T}\left(Q + \pi A^\mathsf{T}A + \pi\widetilde{G}^\mathsf{T}\widetilde{G}\right)x - \left(b + \pi A^\mathsf{T}w + \pi\widetilde{G}^\mathsf{T}\tilde{u}\right)^\mathsf{T}x \\ &\qquad + c^\mathsf{T}z + \pi\|w\|_2^2 + \pi\|\tilde{u}\|_2^2 \\ &z_i = 0 \Rightarrow x_i = 0 \qquad \forall i = 1,\dots,n \\ &z \in \mathscr{O} \\ &z \in \{0,1\}^n,\ x \in \mathbb{R}^n,\ \eta \in \mathbb{R} \end{aligned}$$

yields inequality (11), which is also valid for (12). With the same argumentation as in the first case we have that (11) is valid for F. □

Our intention is to have a cut independent of π. Hence, we analyze the behavior for $\pi \to \infty$. For this, let us denote the solution of

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & \phi(x,z) \\ \text{s.t.} \quad & A(z\circ x) = w \\ & G(z\circ x) \geq u \end{aligned} \tag{13}$$

by $\hat{x}^{\text{con}}(z)$. Furthermore, we denote the corresponding Lagrange multipliers belonging to the equalities by $\mu^{\text{con}}(z)$ and the Lagrange multipliers belonging to the inequalities by $\lambda^{\text{con}}(z)$. Before considering the limit of π we need to make sure that the set $V(\hat{x}_\pi^{\text{pen}}(z),z)$ does not change with increasing π.

Lemma 4 *The set $V(\hat{x}_\pi^{pen}(z), z)$ is independent of π, i.e., changing π does not alter the set.*

The proof of Lemma 4 is stated in Appendix C. We can now examine the behavior for $\pi \to \infty$. We define

$$\delta_i^{\text{con}}(z) \coloneqq \frac{(\hat{x}^{\text{con}}(\sqrt{z})_i)^2}{z_i},$$
$$\gamma_i^{\text{con}}(z) \coloneqq \left(b_i - q_i^\mathsf{T}\left(\sqrt{z} \circ \hat{x}^{\text{con}}(\sqrt{z})\right) + a_i^\mathsf{T}\mu^{\text{con}}(\sqrt{z}) + g_i^\mathsf{T}\lambda^{\text{con}}(\sqrt{z})\right)^2$$

and can derive a cut relying on the Lagrange multipliers instead of the parameter π.

Theorem 6 *Let $\bar{z} \in \mathscr{O} \cap [0,1]^n$ such that $\hat{x}^{con}(\sqrt{\bar{z}})$ exists. Then, the inequality*

$$\eta \geq \phi\left(\hat{x}^{con}(\sqrt{\bar{z}}), \sqrt{\bar{z}}\right) - \frac{1}{2}\sum_{i \in I(\bar{z})}^{n} d_i \delta_i^{con}(\bar{z})(z_i - \bar{z}_i) - \frac{1}{2}\sum_{j \in J(\bar{z})}^{n} \frac{\gamma_j^{con}(\bar{z})}{d_j} z_j + c^\mathsf{T}(z + \sqrt{\bar{z}}) \tag{14}$$

is valid for ($\text{QIND}_{\text{feas}}$).

Proof We use Lemma 3 and show that for $\pi \to \infty$ the inequality (11) converges to inequality (14). We first reiterate an important results concerning penalized optimization (Nocedal and Wright, 2006): the optimal value and solution of (10) converge to the optimal value and solution of (13), i.e., we have

$$\lim_{\pi \to \infty} \hat{x}_\pi^{\text{pen}}(\bar{z}) = \hat{x}^{\text{con}}(\bar{z}), \tag{15}$$
$$\lim_{\pi \to \infty} \phi_\pi^{\text{pen}}(\hat{x}_\pi^{\text{pen}}(\bar{z}), \bar{z}) = \phi(\hat{x}^{\text{con}}(\bar{z}), \bar{z}). \tag{16}$$

Furthermore, a classical result of the theory of penalization methods is that the equality constraints multiplied by the penalization parameter converge to the respective Lagrange multiplier:

$$\lim_{\pi \to \infty} \pi\left(A\left(\bar{z} \circ \hat{x}_\pi^{\text{pen}}(\bar{z})\right) - w\right) = -\mu^{\text{con}}(\bar{z}). \tag{17}$$

Since Lemma 4 tells us that the set of violated constraints is independent of π, we can omit the max term in the penalization for the respective inequalities and can consider them as quadratic penalization terms. Hence, we also have that

$$\lim_{\pi \to \infty} \pi\left(\left(G\left(\bar{z} \circ \hat{x}_\pi^{\text{pen}}(\bar{z})\right)\right)_j - u_j\right) = -\lambda^{\text{con}}(z) \tag{18}$$

for all $j \in V(\hat{x}_\pi^{\text{pen}}(\bar{z}), \bar{z})$. Since the set of violated inequalities is independent of π, we may simply write $V(z)$ instead of $V(\hat{x}_\pi^{\text{pen}}(\bar{z}), \bar{z})$. Next, we take the limit of inequality (11) with respect to π. It is easy to see that $\lim_{\pi \to \infty} \delta_{\pi,i}^{\text{pen}}(\bar{z}) = \delta_i^{\text{con}}(\bar{z})$ for any $i \in I(\bar{z})$. More consideration is required when taking the limit of $\gamma_{\pi,i}^{\text{pen}}(\bar{z})$ for $j \in J(\bar{z})$. First we arrange $\gamma_{\pi,i}^{\text{pen}}(\bar{z})$ into three parts

$$\rho_\pi^1(z) \coloneqq b_i - q_i^\mathsf{T}\left(\sqrt{z} \circ \hat{x}_\pi^{\text{pen}}(\sqrt{z})\right),$$
$$\rho_\pi^2(z) \coloneqq \pi a_i^\mathsf{T}\left(A\left(\sqrt{z} \circ \hat{x}_\pi^{\text{pen}}(\sqrt{z})\right) - w\right),$$
$$\rho_\pi^3(z) \coloneqq \pi g_i(\hat{x}_\pi^{\text{pen}}(\sqrt{z}), \sqrt{z})^\mathsf{T}\left(G(\hat{x}_\pi^{\text{pen}}(\sqrt{z}), \sqrt{z})\left(\sqrt{z} \circ \hat{x}_\pi^{\text{pen}}(\sqrt{z})\right) - u(\hat{x}_\pi^{\text{pen}}(\sqrt{z}), \sqrt{z})\right)$$

and have that $\gamma_{\pi,i}^{\text{pen}}(\bar{z}) = \left(\rho_\pi^1(\bar{z}) - \rho_\pi^2(\bar{z}) - \rho_\pi^3(\bar{z})\right)^2$. Now, it is easy to see that

$$\lim_{\pi \to \infty} \rho_\pi^1(\bar{z}) = b_i - q_i^\mathsf{T}\left(\sqrt{\bar{z}} \circ \hat{x}^{\text{con}}(\sqrt{\bar{z}})\right),$$
$$\lim_{\pi \to \infty} \rho_\pi^2(\bar{z}) = -a_i^\mathsf{T}\mu^{\text{con}}(\sqrt{\bar{z}}),$$

due to (15) and (17). For $\rho_\pi^3(\bar{z})$ we have that $g_i(\hat{x}_\pi^{\text{pen}}(\sqrt{z}), \sqrt{z})$, $G(\hat{x}_\pi^{\text{pen}}(\sqrt{z}), \sqrt{z})$, and $u(\hat{x}_\pi^{\text{pen}}(\sqrt{z}), \sqrt{z})$ are constant in π due to Lemma 4. Hence, we can write

$$\rho_\pi^3(\bar{z}) = \pi \sum_{j \in V(\sqrt{\bar{z}})} (g_i)_j \left(G\left(\sqrt{\bar{z}} \circ \hat{x}_\pi^{\text{pen}}(\sqrt{\bar{z}})\right) - u\right)_j$$

and with (18) we get

$$\lim_{\pi \to \infty} \rho_\pi^3(\bar{z}) = \sum_{j \in V(\sqrt{\bar{z}})} -(g_i)_j(\lambda^{\text{con}}(\sqrt{\bar{z}}))_j.$$

Now we wish to extend the sum to iterate over all $0 \leq j \leq l$. For that we show that

$$-(g_i)_j(\lambda^{\text{con}}(\sqrt{\bar{z}}))_j = 0$$

for all $j \in \{1,\dots,l\} \setminus V(\sqrt{\bar{z}})$. Let us first consider all

$$j \in \{k \in \{1,\dots,l\} : \left(G\hat{x}_\pi^{\text{pen}}(\sqrt{\bar{z}}) - u\right)_k = 0\} =: E, \tag{19}$$

that is, all the the inequalities where equality holds for the optimal point of the penalized problem given a fixed π. That means, we could remove such inequalities and the optimal point would remain optimal. If this was not the case there would be a penalization parameter π sufficiently small where said inequalities would be violated. This however, would contradict Lemma 4. Hence, E is independent of π as well. Since the considered inequalities are redundant to the penalized problem, they are also redundant for the original restricted problem. With standard results from sensitivity analysis we get that $(\lambda^{\text{con}}(\sqrt{\bar{z}}))_j$ must be zero for all $j \in E$.

Next, let us consider all

$$j \in \left\{k \in \{1,\dots,l\} : \left(G\hat{x}_\pi^{\text{pen}}(\sqrt{\bar{z}}) - u\right)_k > 0\right\} =: F$$

where once again F can be computed for a fixed π but ultimately is independent of the penalization parameter. Moreover, all inequalities indexed by F are inactive for $\hat{x}^{\text{con}}(\sqrt{\bar{z}})$ as well. Since $\hat{x}^{\text{con}}(\sqrt{\bar{z}})$ is an optimal point we know that the complementary condition must hold, i.e., for all $j \in F$ it holds that $(\lambda^{\text{con}}(\sqrt{\bar{z}}))_j = 0$. All in all we have that

$$\lim_{\pi\to\infty} \rho_\pi^3(\bar{z}) = \sum_{j=1}^{l} -(g_i)_j(\lambda^{\text{con}}(\sqrt{\bar{z}}))_j = -g_i^\mathsf{T}\lambda^{\text{con}}(\sqrt{\bar{z}}).$$

With that we have all the ingredients to conclude that

$$\lim_{\pi\to\infty} \gamma_{\pi,i}^{\text{pen}}(\bar{z}) = \gamma_i^{\text{con}}(\sqrt{\bar{z}}).$$

Applying the limit (16) proves the proposition. □

The theorem provides the main result. Computing a cut only requires the primal solution $\hat{x}^{\text{con}}(\sqrt{\bar{z}})$ and the dual solution $\left(\mu^{\text{con}}(\sqrt{\bar{z}}), \lambda^{\text{con}}(\sqrt{\bar{z}})\right)$ of a standard QP with equality and inequality constraints. Hence, we can efficiently compute the inequality (14), which provides an effective cut for the original problem (QIND). In Section 6.1 we also handle the cases for which $\hat{x}^{\text{con}}(\sqrt{\bar{z}})$ does not exist.

6 Computational study

In this section we assess the proposed cutting planes empirically via a computational study. For this task, we will utilize an outer approximation, i.e., instead of using the cuts in addition to the original formulation we give the solver no information about the feasible region and only add a cut when a found binary point is infeasible. We describe the outer approximation in detail in the next section. We then study the behavior of the approach on the separable quadratic uncapacitated facility location problem, which is explained in detail in Section 6.2.

6.1 Outer Approximation

The valid inequalities discussed above can be incorporated in a cutting plane algorithm using the outer approximation framework introduced by Duran and Grossmann (1986). Alongside other multi-tree methods based on the generalized Benders decomposition (Geoffrion, 1972) or extended cutting planes (Westerlund and Pettersson, 1995), outer approximation is one of the well-established solution techniques for MINLP (Lee and Leyffer, 2012) and allows for powerful implementations (Abhishek et al., 2010). Recently, algorithms of the outer approximation type have been applied successfully to quadratic integer programs by Bertsimas and Cory-Wright (2021) and Bertsimas and van Parys (2020).

We formulate the outer approximation in Algorithm 1 specifically for the setting discussed above and refer to Li and Sun (2006, Chapter 13) for a general introduction to the method and related approaches for MINLP. The particularity of the approach in this setting is that we only have to compute cuts at the visited binary points. That is, only finitely many cutting planes must be computed – in the worst case for each $z \in \mathcal{O} \cap \{0,1\}$ one inequality has to be computed. So the approach is guaranteed to terminate. It can however happen that we consider a binary point $\bar{z}$ for which there is no feasible x. In this case, a feasibility cut of the following form

$$z^\top \bar{z} <= \bar{z}^\top \bar{z} - 1 \tag{20}$$

Input: Instance $\mathscr{I}$ of (QIND), initial solution z^0 of $\mathscr{I}$
Output: Solution (x^*, z^*) of $\mathscr{I}$

$z^* \leftarrow z^0$;
Let C be the feasible region starting with $C = \mathbb{R}^{n+1}$;
Compute (14) at z^* and cut C by the inequality;
repeat
  Solve problem (OA):
    $\min_z \ \eta$
    s.t. $(\eta, z) \in C$,
      $z \in \{0,1\} \cap \mathscr{O}$.
  Let z^* be the solution of (OA);
  if *the set* $\{x \in \mathbb{R}^n : A(z^* \circ x) = w,\ G(z^* \circ x) \geq u\}$ *is non-empty* **then**
    $x^* \leftarrow \hat{x}^{\text{con}}(\sqrt{z^*})$;
    Compute (14) at z^* and cut C by the inequality;
  else
    Add the inequality (20) at the point z^* to C;
until z^* *is an optimal solution of (OA)*
return (x^*, z^*);

Algorithm 1: Outer approximation for (QIND).

is added, such that we do not visit $\bar{z}$ again.

Most modern MIP solvers have support for handling lazy constraints. Introduced by Barnhart et al. (1998), lazy constraints are constraints which are given to the solver "just in time". That means, the solver is not aware of certain constraints and is only notified of them when a solution is found which violates them. This feature makes implementing an outer approximation not only simple but also efficient. Instead of rebuilding the optimization problem (OA) each time, a lazy constraint is provided. Then the solver terminates if no new lazy constraint is given and the best solution is certified to be optimal.

6.2 Study setting

We apply our method to the so called separable quadratic uncapacitated facility location problem. Given a set of customers J and a set of facilities I, the objective of the problem is to decide which facilities to open, and if, how much they should supply individual customers. Opening a facility $i \in I$ costs a fixed amount of money c_i and the cost of supplying a customer is based on the distance from the facility. Supply and demand is handled proportionally, that means each facility can send a certain percentage to a customer and each customer must get 100% of the deliveries no matter from which facility.

In detail, we are looking at the optimization problem

$$\begin{aligned} \min \quad & \sum_{i\in I} c_i z_i + \sum_{i\in I}\sum_{j\in J} q_{ij}^2 x_{ij} \\ \text{s.t.} \quad & x_{ij} \leq z_i \quad \forall i \in I, j \in J, \\ & \sum_{i\in I} x_{ij} = 1 \quad \forall j \in J, \\ & x_{ij} \geq 0 \quad \forall i \in I, j \in J, \\ & z \in \{0,1\}^I. \end{aligned} \tag{SQUFL}$$

The problem was first studied by Günlük et al. (2007) and later investigated by Günlük and Linderoth (2012) in terms of the perspective reformulation. In fact, the problem is a well suited candidate for the perspective reformulation and hence, fits into the framework presented here. We can leverage the separable structure of the objective function and choose $Q = D$, which yields a strong relaxation.

6.3 Considered instances and methods

For the experiments we consider the same setup as Günlük and Linderoth (2012) and Günlük et al. (2007). That is, we randomly draw the locations p_i and p_j uniformly from $[0,1]^2$ for all $i \in I$ and $j \in J$. Next, we choose $q_{ij} = 50\|p_i - p_j\|_2$. Finally, we draw c_i uniformly from the interval $[1, 100]$.

We consider six different settings. First, we combine $|I| \in \{20, 30\}$ with both $|J| \in \{100, 150\}$ customers, yielding four different setting. Second, we take $|I| = 200$ facilities and combine them with $|J| \in \{200, 400\}$ customers, defining two additional larger settings. Each setting is computed ten times and average runtimes are reported.

We are comparing the outer approximation approach presented in this article, the perspective reformulation (PERSP), and the original logical formulation (QIND).

6.4 Implementation details

The code is implemented in Python and the outer approximation is realized with Gurobi (version 9.1.1). The subproblems for deriving a cut are also solved via Gurobi. It showed that is is beneficial to collect the tested integer points where a lazy constraint is applied and submit the points to Gurobi in a callback as incumbents. As we only have to correct the objective value, i.e., correct η, we immediately regain a feasible point. It seems that Gurobi naturally does not acknowledge this circumstance and must rediscover the points. Submitting visited binary points as incumbents improves the runtime significantly. The perspective reformulation and the logical formulation are solved with Gurobi as well.

The code of the implementation as well as the experimental setup is publicly available, see `https://gitlab.com/Dnis/tango`. All computations were run on a computer with an Intel Core i7-6700 CPU 3.40GHz and 32GB RAM.

6.5 Results

It is well known that the the perspective reformulation works particularly well on the SQUFL problem (Günlük and Linderoth, 2012). Our results show that even though the perspective improves the runtimes considerably compared to the naive logical formulation, the outer approximation enables even faster computation times. Table 1 shows the mean runtimes for the three approaches. We observe that the outer approximation solves the instances the fastest on average, being able to solve the highest dimensional instance about 7 times faster.

Table 1 Mean runtimes for the different SQUFL instances in seconds. Data for the logical formulation for $|I| = 200$ is missing as the problems were not solved within the time limit of 3600 seconds.

Instances				
$\|I\|$	$\|J\|$	Logical formulation	Perspective reformulation	Outer approximation
20	100	57.873	2.081	**1.446**
	150	154.270	3.856	**2.044**
30	100	482.760	3.831	**1.992**
	150	1662.118	6.823	**2.871**
200	200	NA	395.208	**62.128**
200	400	NA	1496.675	**210.038**

Assessing the cumulative distribution of the runtimes in Figures 1 and 2, we can see that the outer approximation approach dominates the perspective reformulation significantly in all cases. Of particular interest is that in large scale experiments we can see that the outer approximation has nearly no variance in the observed runtimes.

We can state that the outer approximation is overall much faster than the other algorithms. Of course, we want to discuss reasons for the promising results next. After all, we are operating on a relaxation which is as tight as the relaxation of the perspective relaxation. Looking at the required number of branch-and-bound nodes gives us further insights, which helps us answer the question. Figure 3 shows a cumulative distribution plot of the percentage of solved instances versus the required number of branch-and-bound nodes. We are omitting the results for the low-dimensional instances as they are not showing any further insights compared to the high-dimensional results. We can see in Figure 3 that generally the outer approximation must visit less nodes than the perspective reformulation. However, the difference is far less pronounced than the runtimes would suggest. We interpret the results as such that the perspective reformulation approach needs to solve a nonlinear second-order cone problem in each node, which on top has much more variables and constraints. For the outer approximation the continuous variables and respective constraints are only considered when we compute a tangent cut. In this cases all binary constraints are fixed. Hence, we only have to solve a possibly low dimensional quadratic program. For the binary proportion of the problem, Gurobi only has to solve a linear binary program with one continuous variable. We conclude that this results in less computational burden, which allows for a much faster solving time.

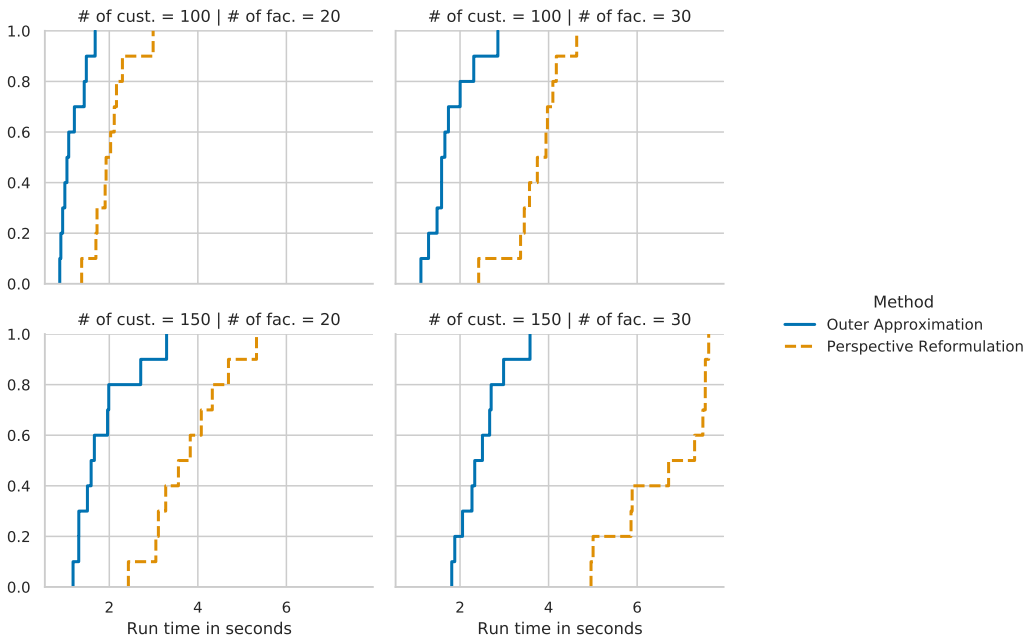

Fig. 1 Percentage of solved instances within the displayed time. The plots show the results for the instances with 20 and 30 facilities and either 100 or 150 customers.

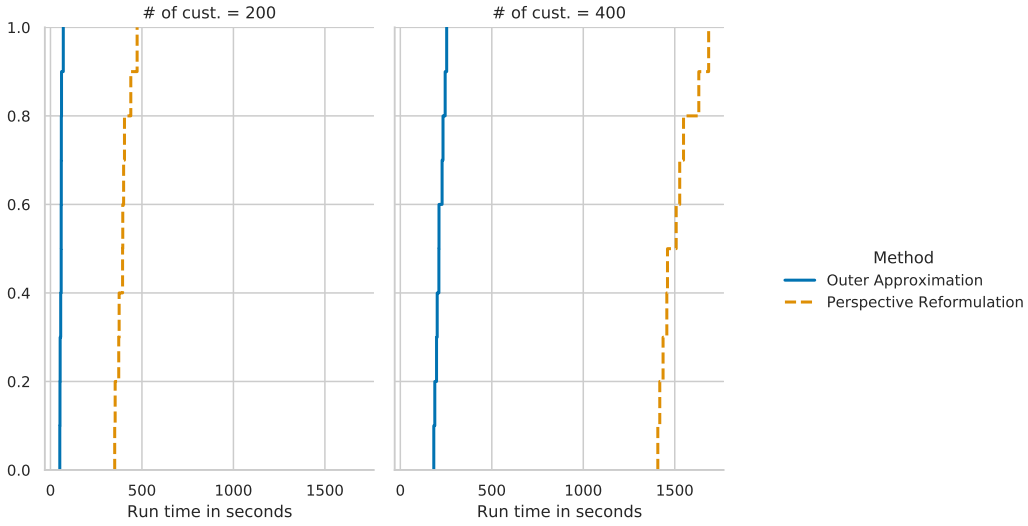

Fig. 2 Percentage of solved instances within the displayed time. The two plots show the results for the instances with 200 facilities and either 200 or 400 customers.

7 Conclusion

We have studied a general version (QIND) of the quadratic optimization problem with indicator constraint. Our analysis was based on a sequence of relaxations that are inspired by previous results on optimization problems with indicator constraints – most prominently the perspective reformulation (PERSP). We introduced the new relaxation (CMF) and proved that it is as strong as (PERSP).

Since the perspective reformulation is known to be a strong relaxation, this structural result justified research on valid inequalities derived from (CMF). We developed projective cuts using a three step technique: First, we

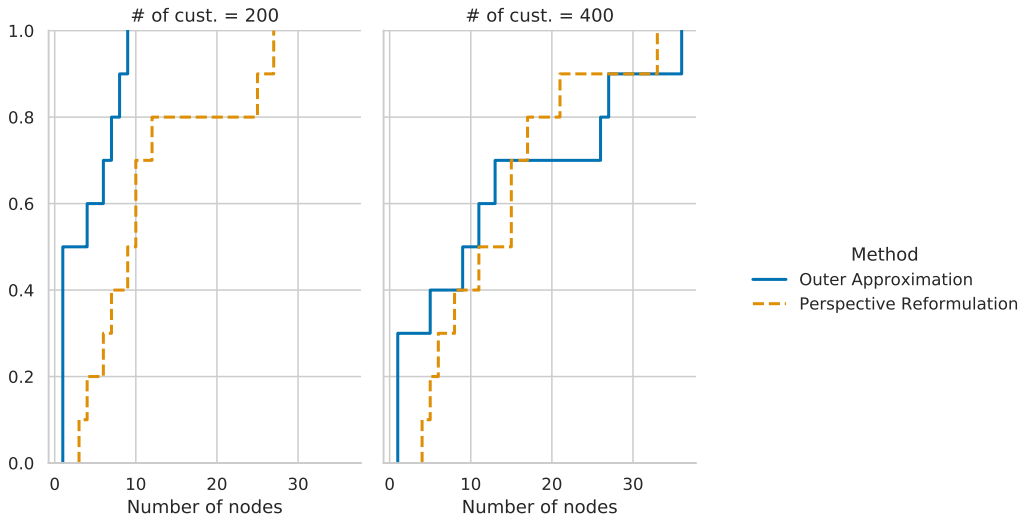

Fig. 3 Percentage of solved instances while only using a maximum number of branch-and-bound nodes. The two plots show the results for the instances with 200 facilities and either 200 or 400 customers.

proved the validity of projective tangent cuts for a ε-disturbed problem without additional inequality and equality constraints. Second, we evaluated the limits for $\varepsilon \to 0$ to obtain cuts without the additional parameter. In the third and last step, we re-introduced the original inequality and equality constraints and showed how they can be incorporated in our framework with the help of a penalized auxiliary problem. The developed theory generalizes recent results on tangent cuts by other authors and for the first time gives all details of the tangent computations in a self-contained presentation.

The computational experiments of Section 6 show that the theory translates into promising algorithmic improvements. In particular, the comparison with the state-of-the-art perspective reformulation shows the high potential of projective cuts. Our Python implementation is not specifically optimized for speed and we expect that we did not exhaust the full computational potential of perspective cuts, leaving space for follow-up research on the implementation details. The code is publicly available to allow for assessment and future experiment, see `https://gitlab.com/Dnis/tango`.

The presented theoretic framework is not strictly limited to the problem formulation (QIND) studied here. Therefore, generalization of the proof techniques detailed in this contribution are certainly possible. In particular, we have identified the extension of the projective cut framework to other MINLP relaxations as a future research direction. In addition, a promising idea for improvement of the algorithm is choosing an optimized matrix D for the computation of each cut instead of fixing D for the entire algorithm.

Acknowledgements The second author is supported within the research training group 2126 Algorithmic Optimization (ALOP) founded by the German Research Foundation DFG.

References

Abhishek K, Leyffer S, Linderoth J (2010) FilMINT: an outer approximation-based solver for convex mixed-integer nonlinear programs. INFORMS Journal on Computing 22(4):555–567

Atamtürk A, Gómez A (2018) Strong formulations for quadratic optimization with M-matrices and indicator variables. Mathematical Programming 170(1):141–176

Barnhart C, Johnson EL, Nemhauser GL, Savelsbergh MWP, Vance PH (1998) Branch-and-price: Column generation for solving huge integer programs. Operations Research 46(3):316–329

Belotti P, Bonami P, Fischetti M, Lodi A, Monaci M, Nogales-Gómez A, Salvagnin D (2016) On handling indicator constraints in mixed integer programming. Computational Optimization and Applications 65(3):545–566

Bertsimas D, Cory-Wright R (2021) A scalable algorithm for sparse portfolio selection. arXiv e-prints https://arxiv.org/abs/1811.00138

Bertsimas D, van Parys B (2020) Sparse high-dimensional regression: exact scalable algorithms and phase transitions. Annals of Statistics 48(1):300–323

Bertsimas D, Shioda R (2009) Algorithm for cardinality-constrained quadratic optimization. Computational Optimization and Applications 43(1):1–22

Bertsimas D, King A, Mazumder R (2016) Best subset selection via a modern optimization lens. The Annals of Statistics 44(2):813–852

Bertsimas D, Cory-Wright R, Pauphilet J (2021) A unified approach to mixed-integer optimization problems with logical constraints. arXiv e-prints https://arxiv.org/abs/1907.02109

Bienstock D (1996) Computational study of a family of mixed-integer quadratic programming problems. Mathematical Programming 74(2):121–140

Bonami P, Lodi A, Tramontani A, Wiese S (2015) On mathematical programming with indicator constraints. Mathematical Programming 151(1):191–223

Burke JV, Hoheisel T (2015) Matrix support functionals for inverse problems, regularization, and learning. SIAM Journal on Optimization 25(2):1135–1159

Dong H, Chen K, Linderoth J (2015) Regularization vs. relaxation: a conic optimization perspective of statistical variable selection. arXiv e-prints https://arxiv.org/abs/1510.06083

Duran MA, Grossmann IE (1986) An outer-approximation algorithm for a class of mixed-integer nonlinear programs. Mathematical Programming 36(3):307–339

Frangioni A, Gentile C (2006) Perspective cuts for a class of convex 0-1 mixed integer programs. Mathematical Programming 106(2):225–236

Frangioni A, Gentile C (2007) SDP diagonalizations and perspective cuts for a class of nonseparable MIQP. Operations Research Letters 35(2):181–185

Frangioni A, Gentile C, Grande E, Pacifici A (2011) Projected perspective reformulations with applications in design problems. Operations Research 59(5):1225–1232

Gao J, Li D (2011) Cardinality constrained linear-quadratic optimal control. IEEE Transactions on Automatic Control 56(8):1936–1941

Geoffrion AM (1972) Generalized benders decomposition. Journal of optimization theory and applications 10(4):237–260

Günlük O, Linderoth J (2010) Perspective reformulations of mixed integer nonlinear programs with indicator variables. Mathematical Programming 124(1-2):183–205

Günlük O, Linderoth J (2012) Perspective reformulation and applications. In: Mixed Integer Nonlinear Programming, Springer, New York, pp 61–89

Günlük O, Lee J, Weismantel R (2007) MINLP strengthening for separable convex quadratic transportation-cost UFL. IBM Research Report pp 1–16

Konno H, Yamamoto R (2009) Choosing the best set of variables in regression analysis using integer programming. Journal of Global Optimization 44:273–282

Kreber D (2019) Cardinality-Constrained Discrete Optimization for Regression. PhD thesis, Trier University

Lee J, Leyffer S (eds) (2012) Mixed Integer Nonlinear Programming. Springer, New York

Li D, Sun X (2006) Nonlinear Integer Programming. Springer, New York

Markowitz H (1952) Portfolio selection. The Journal of Finance 7(1):77–91

Meyer CD (2000) Matrix Analysis and Applied Linear Algebra. SIAM

Miller A (1990) Subset Selection in Regression. Chapman and Hall, Melbourne

Natarajan BK (1995) Sparse approximate solutions to linear systems. SIAM Journal on Computing 24(2):227–234

Nesterov Y, Nemirovskii A (1994) Interior-Point Polynomial Algorithms in Convex Programming. SIAM

Nocedal J, Wright S (2006) Numerical Optimization. Springer Science & Business Media

Wei D, Sestok CK, Oppenheim AV (2013) Sparse filter design under a quadratic constraint: low-complexity algorithms. IEEE Transactions on Signal Processing 61(4):857–870

Westerlund T, Pettersson F (1995) An extended cutting plane method for solving convex minlp problems. Computers & Chemical Engineering 19:131–136

Xie W, Deng X (2020) Scalable algorithms for the sparse ridge regression. SIAM Journal on Optimization 30(4):3359–3386

Zheng X, Sun X, Li D (2014) Improving the Performance of MIQP Solvers for Quadratic Programs with Cardinality and Minimum Threshold Constraints: a Semidefinite Program Approach. INFORMS Journal on Computing 26(4):690–703

Appendix

A Proof of Lemma 1

Let i be an element of $\{1,\dots,n\}$. We consider a solution of the problem

$$\begin{aligned} \min \quad & \|u\|^2 \\ \text{s.t.} \quad & \left(h_\varepsilon^i\right)^\mathsf{T} u = \left(h_\varepsilon^i\right)^\mathsf{T} \hat{u}_i. \end{aligned} \tag{21}$$

Then, $\hat{u}_i$ must be optimal for (21). Otherwise there is an optimal solution $\tilde{u}_i$ such that $\|\tilde{u}_i\|^2 < \|\hat{u}_i\|^2$ and such that $(\hat{\eta}, \hat{\tau}, \hat{u}_0, \dots, \tilde{u}_i, \dots \hat{u}_n)$ is a feasible point. It is easy to see that then $\hat{\tau}_i$ can be chosen smaller and hence $\hat{\eta}$ can be decreased as well. This, however, implies that $(\hat{\eta}, \hat{\tau}, \hat{u}_0, \dots, \hat{u}_n)$ is not optimal, which is a contraction to our assumption.

Due to the KKT conditions and the strict convexity of (21), a vector u is optimal if and only if it solves the system of equations

$$\begin{aligned} 2u + \lambda h_\varepsilon^i &= 0 \\ \left(h_\varepsilon^i\right)^\mathsf{T} u &= \left(h_\varepsilon^i\right)^\mathsf{T} \hat{u}_i, \end{aligned}$$

where $\lambda \in \mathbb{R}$ is the Lagrange multiplier. From the system it is easy to see that any optimal point has to be a multiple of h_ε^i, hence for every i the set $\{\hat{u}_i, h_\varepsilon^i\}$ is linear dependent.

B Proof of Lemma 2

We first show statement i). Using the continuity of matrix inversion, it is easy to see that

$$\begin{aligned} \lim_{\varepsilon\to 0} \hat{x}_\varepsilon(z) &= \lim_{\varepsilon\to 0} \left(H_\varepsilon(z)^\mathsf{T} H_\varepsilon(z) + D_\varepsilon\right)^{-1} b(z) \\ &= \lim_{\varepsilon\to 0} \left(\operatorname{diag}(z)\,(Q - D_\varepsilon)\operatorname{diag}(z) + D_\varepsilon\right)^{-1} b(z) \\ &= \lim_{\varepsilon\to 0} \left(\operatorname{diag}(z)\,(Q - D + \varepsilon I)\operatorname{diag}(z) + D - \varepsilon I\right)^{-1} b(z) \\ &= \lim_{\varepsilon\to 0} \left(\operatorname{diag}(z)\,(Q - D)\operatorname{diag}(z) + D + \varepsilon\left(\operatorname{diag}(z)^2 - I\right)\right)^{-1} b(z) \\ &= \lim_{\varepsilon\to 0} \left(H(z)^\mathsf{T} H(z) + D + \varepsilon\left(\operatorname{diag}(z)^2 - I\right)\right)^{-1} b(z) \\ &= \left(H(z)^\mathsf{T} H(z) + D + \lim_{\varepsilon\to 0} \varepsilon\left(\operatorname{diag}(z)^2 - I\right)\right)^{-1} b(z) \\ &= \left(H(z)^\mathsf{T} H(z) + D\right)^{-1} b(z) \\ &= \hat{x}(z) \end{aligned}$$

and hence the statement follows. With the limit of $\hat{x}_\varepsilon$ we can determine the limit of $\nabla\zeta_\varepsilon$ and afterwards the limit of ζ_ε.

From Theorem 3 and i) it follows that

$$\begin{aligned} & \lim_{\varepsilon\to 0} \frac{\partial}{\partial z_i} \frac{1}{2} y_\varepsilon^\mathsf{T} \left(I + \sum_{i=1}^n \frac{1}{d_i - \varepsilon} h_\varepsilon^i \left(h_\varepsilon^i\right)^\mathsf{T} z_i\right)^{-1} y_\varepsilon \\ = & \lim_{\varepsilon\to 0} -\frac{\left(b_i - \left(h_\varepsilon^i\right)^\mathsf{T} H_\varepsilon(\sqrt{z}) \cdot \hat{x}_\varepsilon(\sqrt{z})\right)^2}{2(d_i - \varepsilon)} \\ = & \lim_{\varepsilon\to 0} -\frac{\left(b_i - \left(h_\varepsilon^i\right)^\mathsf{T} H_\varepsilon \cdot (\sqrt{z} \circ \hat{x}_\varepsilon(\sqrt{z}))\right)^2}{2(d_i - \varepsilon)} \\ = & \lim_{\varepsilon\to 0} -\frac{\left(b_i - (q_i - d_i e_i + \varepsilon e_i)^\mathsf{T} (\sqrt{z} \circ \hat{x}_\varepsilon(\sqrt{z}))\right)^2}{2(d_i - \varepsilon)} \\ = & -\frac{\left(b_i - (q_i - d_i e_i)^\mathsf{T} (\sqrt{z} \circ \hat{x}(\sqrt{z}))\right)^2}{2d_i}. \end{aligned}$$

Applying i) to the identity in Theorem 3 yields

$$\lim_{\varepsilon\to 0} -\frac{d_i - \varepsilon}{z_i} \left(\hat{x}_\varepsilon(\sqrt{z})_i\right)^2 = -\frac{d_i}{z_i} \left(\hat{x}(\sqrt{z})_i\right)^2,$$

and thus, the identity for $z_i > 0$ follows. Now assume that $z_i = 0$. Then, we have that

$$d_i e_i^\mathsf{T} \operatorname{diag}\left(\sqrt{z}\right) \hat{x}\left(\sqrt{z}\right) = d_i \sqrt{z_i} \hat{x}\left(\sqrt{z}\right)_i = 0.$$

It holds that

$$\begin{aligned}\lim_{\varepsilon\to 0}\zeta_\varepsilon(z) &= \lim_{\varepsilon\to 0}\psi_\varepsilon(\sqrt{z})+c^\mathsf{T}\left(z-\sqrt{z}\right)\\ &= \lim_{\varepsilon\to 0}\min_{x\in\mathbb{R}^n}\phi_\varepsilon(x,\sqrt{z})+c^\mathsf{T}\left(z-\sqrt{z}\right)\\ &= \lim_{\varepsilon\to 0}\phi_\varepsilon(\hat{x}_\varepsilon(\sqrt{z}),\sqrt{z})+c^\mathsf{T}\left(z-\sqrt{z}\right)\\ &= \lim_{\varepsilon\to 0}\frac{1}{2}\hat{x}_\varepsilon(\sqrt{z})^\mathsf{T}H_\varepsilon(\sqrt{z})^\mathsf{T}H_\varepsilon(\sqrt{z})\hat{x}_\varepsilon(\sqrt{z})\\ &\qquad +\frac{1}{2}\hat{x}_\varepsilon(\sqrt{z})^\mathsf{T}D_\varepsilon\hat{x}_\varepsilon(\sqrt{z})-b^\mathsf{T}\hat{x}_\varepsilon(\sqrt{z})+c^\mathsf{T}z.\end{aligned}$$

Letting ε converge to 0 and utilizing i) yields

$$\begin{aligned}&\frac{1}{2}\hat{x}(\sqrt{z})^\mathsf{T}H(\sqrt{z})^\mathsf{T}H(\sqrt{z})\hat{x}(\sqrt{z})+\frac{1}{2}\hat{x}(\sqrt{z})^\mathsf{T}D\hat{x}(\sqrt{z})-b^\mathsf{T}\hat{x}(\sqrt{z})+c^\mathsf{T}z\\ =\;&\phi(\hat{x}(\sqrt{z}),\sqrt{z})+c^\mathsf{T}\left(z-\sqrt{z}\right),\end{aligned}$$

which proves the lemma.

C Proof of Lemma 4

We are proving the statement in a general setting as it does not require all the extra notations we introduced. We consider the problem

$$\begin{aligned}\min_x\quad & f(x)\\ \text{s.t.}\quad & q(x)=0\\ & p(x)\geq 0\end{aligned}$$

where $f:\mathbb{R}^n\to\mathbb{R}$, $q:\mathbb{R}^n\to\mathbb{R}^m$, and $p:\mathbb{R}^n\to\mathbb{R}^l$ are continuously differentiable and convex functions. The penalized problem is then given as

$$\min_x f(x)+\pi\sum_{i=1}^m q_i^2(x)+\pi\sum_{i=1}^l\max(0,p_i^2(x)). \tag{22}$$

We denote the solution to (22) by $\hat{x}_\pi$, we denote the set of violated inequalities by V_π, i.e.,

$$V_\pi := \{i\in\{1,\dots,l\} : p_i(\hat{x}_\pi)<0\},$$

and we define the set

$$A_\pi := \{i\in\{1,\dots,l\} : p_i(\hat{x}_\pi)\leq 0\}.$$

Now assume that the proposed statement is false, that is, there are $\pi_0,\pi_1>0$ such that $V_{\pi_0}\supset V_{\pi_1}$. Moreover, without loss of generality we assume that π_1 is chosen in a way such that for the set

$$W := V_{\pi_0}\setminus V_{\pi_1}$$

it holds that

$$p_i(\hat{x}_{\pi_1})=0$$

for all $i\in W$. Note that $\hat{x}_\pi$ is also an optimal point of

$$\min_x\; r(x,A_\pi) := f(x)+\pi\sum_{i=1}^m q_i^2(x)+\pi\sum_{i\in A_\pi}p_i^2(x).$$

We then have $r(x,A_{\pi_0})=r(x,A_{\pi_1})+\pi\sum_{i\in W}p_i^2(x)$. Computing the gradient of r_{π_1} at the point $\hat{x}_{\pi_1}$ yields

$$\nabla_x r(\hat{x}_{\pi_1},A_{\pi_1})=\nabla_x r(\hat{x}_{\pi_1},A_{\pi_0})-2\pi\sum_{i\in W}p_i(\hat{x}_{\pi_1})\nabla_x p_i(\hat{x}_{\pi_1}).$$

Since $p_i(\hat{x}_{\pi_1})=0$ holds for all $i\in W$, it follows that

$$\nabla_x r(\hat{x}_{\pi_1},A_{\pi_1})=\nabla_x r(\hat{x}_{\pi_1},A_{\pi_0}).$$

However, by assumption, $\hat{x}_{\pi_1}$ is not an optimal point of $\min_x r(x,A_{\pi_0})$, and hence, the gradient is not zero. That implies that $\nabla_x r(\hat{x}_{\pi_1},A_{\pi_1})$ is not equal to the zero vector either, which means that $\hat{x}_{\pi_1}$ is not an optimal point of $\min\; r(x,A_{\pi_0})$ – a contradiction to our assumption.

Thus, the sets of violated inequalities cannot change depending on the choice of the penalization parameter π. Applying this general statement to the setting in this article yields the desired result.