Dantzig-Wolfe Reformulations for Binary Quadratic Problems

Alberto Ceselli[1]*, Lucas Létocart[2], Emiliano Traversi[2]

November 26, 2022

1. Università degli Studi di Milano
Dipartimento di Informatica
Via Celoria, 18
20133 Milano, Italy.
alberto.ceselli@unimi.it

2. Université Sorbonne Paris Nord
LIPN, CNRS, (UMR 7030)
99 Avenue Jean Baptiste Clément
F-93430 Villetaneuse, France.
{*lucas.letocart, emiliano.traversi*}*@lipn.univ-paris13.fr*

Abstract

The purpose of this paper is to provide strong reformulations for binary quadratic problems. We propose a first methodological analysis on a family of reformulations combining Dantzig-Wolfe and Quadratic Convex optimization principles. We show that a few reformulations of our family yield continuous relaxations that are strong in terms of dual bounds and computationally efficient to optimize. As a representative case study, we apply them to a cardinality constrained quadratic knapsack problem, providing extensive experimental insights. We report and analyze in depth a particular reformulation providing continuous relaxations whose solutions turn out to be integer optima in all our tests.

1 Introduction

Generic solvers for mathematical programming models have been steadily improving since long time, in terms of both computing capabilities and types of models they can handle. While up to few years ago the focus was on effectively solving Mixed Integer Linear Programs (MILP), the interest is currently shifting to more general problems, possibly coping with nonlinearities in either the objective function or the constraints. Binary Quadratic Problems (BQPs) are among the special classes of

*Partially funded by Regione Lombardia - Fondazione Cariplo, grant n. 2015-0717, project "RED-NEAT", and partially undertook when A. Ceselli was visiting LIPN - Université Sorbonne Paris Nord.

Mixed Integer Non-Linear Problems (MINLP) which are currently subject to wider investigation.

In its generic form a BQP reads as follows:

$$\text{(BQP)} \quad \min\{x^\top Sx + L^\top x | Ax \geq b, x \in \{0,1\}^n\} \, . \tag{1}$$

with $S \in \mathcal{Q}^{n\times n}$, not restricted to be positive semi-definite, and $L \in \mathcal{Q}^n$, $A \in \mathcal{Q}^{m\times n}$, $b \in \mathcal{Q}^m$. Several generic solvers are already available to tackle BQPs. To mention just a few examples BiqCrunch [25], CPLEX [23], GloMIQO [32], Gurobi [21] and SCIP [37] support direct optimization of either BQPs or even MINLPs. In the vast majority of the cases, these solvers rely on Branch-and-Bound techniques, embedding special relaxations of the models provided by the user.

Restricting to the BQP case, a first popular choice for such relaxations is to reformulate the problem by linearizing the quadratic terms [1, 17, 19, 20, 31, 38], thereby obtaining a MILP. A second promising way to obtain dual bounds for BQPs is to rely on Semidefinite Programming (SDP) relaxations, like for example [25, 36].

When the objective function is convex, even the simple continuous relaxation of a BQP is appealing, as it yields convex optimization subproblems, for which effective algorithms are engineered in state-of-the-art solvers. A third alternative for generic BQPs is therefore to use a reformulation obtained via the so-called *convexification* of the quadratic objective function. This convexification can be obtained by applying Quadratic Convex Reformulation (QCR) methods like those introduced in [8] and extended in [7].

On the other hand, Dantzig-Wolfe Reformulation (DWR) is a well known technique used to obtain tight bounds for MILPs (see for example [13, 14, 39]). Its principle is to replace the feasibility region corresponding to a subset of the constraints of a model by the convex hull of its extreme points through an inner representation. A recent research trend is indeed investigating on how to embed automatic decomposition techniques into general purpose MILP solvers [6, 18, 5].

DWR can be applied in principle also to nonlinear mathematical programming models, provided a subset of constraints exists whose corresponding feasible region can be represented as a polyhedron. In fact, the extension of DWR to nonlinear problems has been analyzed in several theoretical papers in the past years (see for example [2, 22, 26]). However, whether or not its application to MINLP may yield successful computational methods is still an open research question.

One of the main issues is the following: the application of DWR leads to a formulation with an exponential number of variables. It can be solved via iterative procedures like Column Generation (we refer the reader to [14] for an extensive review of such a method), but additional conditions have to be fulfilled to ensure convergence.

Main contributions. The first contribution of this paper is an analysis of how to generalize DWR to tackle 0-1 nonlinear problems (Section 3). As second contribution we present a methodology to systematically combine DWR and convexification methods to effectively solve BQPs (Section 4). As a final contribution we show that applying DWR to BQPs is not only theoretically possible, but also computationally profitable: we consider as test-bed the 0-1 exact k-item Quadratic Knapsack Problem (kQKP), a particular BQP in which the objective function is generic, and the set of constraints is composed by a single inequality and a single equation (Section 5). Remarkably, on that particular problem we obtain a reformulation producing models whose continuous relaxation has an optimal integer solution in all the experiments we have performed. Therefore, we undertake and discuss an additional set of tests in order to better understand such impressive result (Section 6).

2 Basics

In this Section we discuss the concept of Lagrangian Duality and QCR method introducing a unified notation.

2.1 Lagrangian Duality

We refer the reader to [27, 28] for an extensive analysis of Lagrangian Duality in the context of Combinatorial Optimization. Let us consider the following (primal) problem:

$$\begin{aligned} \min \quad & f(x) & \\ \text{s.t.} \quad & g_i(x) \le 0 & i \in I \\ & h_j(x) \le 0 & j \in J \\ & x \in \mathbb{R}^n & \end{aligned}$$

The *Lagrangian Function* $L(x, \mu)$ of the primal variables x and the dual variables (or multipliers) $\mu \in \mathbb{R}_+^{|I|}$ is the following:

$$L(x, \mu) = f(x) + \mu^\top g(x) \ .$$

The definition of the dual variables can be easily extended to the case where for some $i' \in I$ we have equations instead of inequalities by unrestricting the sign of the corresponding multipliers. The *dual function* $\theta(\mu)$ is defined as follows:

$$\theta(\mu) = \min_{\substack{x \in \mathbb{R}^n:\ h_j(x) \le 0 \\ j \in J}} L(x, \mu)$$

and the corresponding *Lagrangian Dual* problem is the following:

$$\max_{\mu \in \mathbb{R}_+^{|I|}} \theta(\mu) = \max_{\mu \in \mathbb{R}_+^{|I|}} \min_{\substack{x \in \mathbb{R}^n:\ h_j(x) \le 0 \\ j \in J}} L(x, \mu).$$

We say that the Lagrangian Function and the associated Lagrangian Dual are obtained by *relaxing* (or *dualizing*) the constraints $g_i(x) \le 0$. Several Lagrangian Duals can be obtained after relaxing different sets of constraints. A key property of the Lagrangian Dual is that (regardless of which set of constraints is relaxed) its optimal solution always gives a valid lower (or dual) bound for the primal problem. It is important to notice that the dual function is *always* concave; therefore the Lagrangian Dual is always convex in the μ multipliers. That is, in principle, it could always be solved to global optimality using convex optimization techniques, provided that, for a fixed μ, evaluating the value of $\theta(\mu)$ can be done effectively.

2.2 Quadratic Convex Reformulation Method

If the problem we want to solve contains binary variables and a quadratic objective function (as in the case of (BQP)), we can reformulate the objective function by exploiting the facts that, on its feasible region, we always have

$$x_j^2 = x_j \quad \forall j = 1 \dots n \tag{2}$$

(because the variables are binary) and, in case we have a set of valid equations for our problem, say $A_= x = b_=$, we also have:

$$(A_= x - b_=)^2 = 0 \tag{3}$$

where, given a generic vector v, we use the notation $v^2 = v^\top v$.

This idea is exploited in the Quadratic Convex Reformulation (QCR) method and extensions (see [8, 7, 35]). By Lagrangian relaxation of constraints (2) and (3) with the multipliers $\delta_j \in \mathbb{R}$ $(j = 1 \dots n)$ and $\rho \in \mathbb{R}$ respectively, we obtain the following Lagrangian Function:

$$L(x, \delta, \rho) = f_{\delta,\rho}(x) = f(x) + \sum_{j \in J} \delta_j (x_j^2 - x_j) + \rho (A_= x - b_=)^2$$

and the corresponding Lagrangian Dual becomes

$$\max_{\delta \in \mathbb{R}^n, \rho \in \mathbb{R}} \theta(\delta, \rho) = \max_{\delta \in \mathbb{R}^n, \rho \in \mathbb{R}} \min_{x \in \mathbb{R}^n : Ax \geq b} f_{\delta,\rho}(x) \ . \tag{4}$$

In [16], the authors show that, in this special case, the Lagrangian Dual reduces to a Semi Definite Program (SDP), and the set of optimal multipliers (δ^*, ρ^*) can be obtained by solving an auxiliary semidefinite problem. The optimal multipliers are selected in a way that the function $f_{\delta^*,\rho^*}(x)$ is convex. This implies that we can obtain a convex objective function by replacing $f(x)$ with $f_{\delta^*,\rho^*}(x)$, in such a way that it can be efficiently solved by a classical branch-and-bound algorithm based on its continuous relaxation. Moreover, among all the possible settings of (δ, ρ) making $f_{\delta,\rho}(x)$ convex, the optimal multipliers (δ^*, ρ^*) are those yielding the tightest possible continuous relaxation.

In [7], the authors propose an improvement of the QCR method based on an extended formulation using an additional set of variables z_{ij}, representing the products of the binary variables x_i and x_j. This allows to have the following, more general, Lagrangian Function:

$$f_{\delta,\rho,\Gamma}(x, z) = f(x) + \sum_{j \in J} \delta_j (x_j^2 - x_j) + \rho (A_= x - b_=)^2 + \sum_{i \in J} \sum_{j \in J} \Gamma_{ij} (z_{ij} - x_i x_j)$$

with $\Gamma \in \mathbb{R}^{n \times n}$. To this formulation are also added the classical linearization constraints introduced in [17]. This additional degree of freedom allows to obtain formulations that are stronger than the one obtained with $f_{\delta,\rho}(x)$ (see [7]). Also in this case, the problem of finding the best set of multipliers $(\delta^*, \rho^*, \Gamma^*)$ reduces to solve an SDP.

For the sake of completeness, in Appendix A we report the two SDP needed to compute the optimal set of multipliers (δ^*, ρ^*) and $(\delta^*, \rho^*, \Gamma^*)$.

3 Dantzig-Wolfe Reformulation for 0-1 nonlinear problems

A popular technique to obtain tight bounds for problems with discrete variables is Dantzig-Wolfe Reformulation (DWR) (see for example [14] and [24], Chapter 13). The main idea behind DWR is to replace the feasible region corresponding to a subset of constraints with the convex hull of its extreme points, exploiting the Minkovsky and Weyl theorem. From an algebraic point of view, this amounts to force the vector of decision variables to be represented as a linear convex combination of a finite set of (extreme) points. For this reason, in the literature this procedure is commonly known as partial "convexification"; however, in order to avoid misunderstanding with the QCR methods that deal with the convexification of the objective function, we refer to the effect of DWR as "strengthening" of a subset of constraints.

DWR has been extensively used for solving problems with only linear constraints and linear objective function. However, it has never been applied systematically to generic nonlinear problems. In this section we give an exact procedure to perform DWR on nonlinear problems. More precisely, we want to apply DWR to the following problem:

$$
\begin{array}{llll}
\text{(BF)} & \min & f(x) & \\
& \text{s.t.} & g_i(x) \le 0 & i \in I \\
& & h_j(x) \le 0 & j \in J \\
& & x \in \{0,1\}^n &
\end{array}
\qquad (5)
$$

where $h_j(x), j \in J$ are convex. The basic idea of DWR is to substitute constraints (5) with a set of constraints imposing on x to belong to their convex hull $\Omega = \text{conv}\{x : h_j(x) \le 0 \quad j \in J, x \in \{0,1\}^n\}$. This leads to a reformulation with a finite number of variables if all x are discrete and/or all constraints are linear. However, if some of the constraints $h_j(x)$ are nonlinear and some of the variables are continuous this may no longer be true and a more careful analysis of the problem is needed to guarantee the convergence of the method (if some of the variables are no longer required to be integer, the size of the set of extreme points of the convex hull of nonlinear constraints can be infinite).

An important way to compare reformulations is to evaluate the strength of their continuous relaxations, i.e. when constraints $x \in \{0,1\}^n$ are substituted by $x \in [0,1]^n$. In the following we indicate with (F) the continuous relaxation of (BF), that is (BF) with variable bounds instead of binary restrictions. For convenience, we assume that constraints $x \in [0,1]^n$ are part of the set of constraints $g_i(x) \le 0, i \in I$. This notation and this assumption apply in the following to all the formulations used. Furthermore we say that a formulation is *stronger* than another if no instance exists in which the dual bound produced by the former is looser (and some instance exists in which it is tighter).

DWR with a Master Problem with nonlinear objective function. Let $\mathcal{P}$ be the set of the extreme points of Ω. For each extreme point $p \in \mathcal{P}$, we define x^p as its incidence vector and we introduce a new variable y^p associated to p. With a slight abuse of notation, in the following we will use the notation $\mathcal{P}$ also for the set of y^p variables associated to the set $\mathcal{P}$.

With the given notation, (BF) can be reformulated as follows:

$$
\begin{aligned}
\text{(BF-NLM)} \quad \min \quad & f(x) & & \\
\text{s.t.} \quad & g_i(x) \leq 0 & i \in I \; [\mu_i] & \\
& x - \sum_{p \in \mathcal{P}} x^p y^p = 0 & [\pi] & \qquad (6) \\
& \sum_{p \in \mathcal{P}} y^p - 1 = 0 & [\pi_0] & \\
& y^p \geq 0 & \forall p \in \mathcal{P} & \\
& x \in \{0,1\}^n & , &
\end{aligned}
$$

where μ_i, π and π_0 denote the dual variables of the associated constraints in its continuous relaxation (F-NLM), which is usually indicated as Master Problem. We refer to the reformulation (F-NLM) as *DWR with a nonlinear Master Problem.* The reformulation is valid for any $f(x)$ and $g_i(x)$. It is stronger than (F) when used as a dual bound for (BF). The drawback of the new formulation is that it can contain a very large number of variables (one for each $p \in \mathcal{P}$). The standard technique used in practice to solve such formulation is *Column Generation* (CG). This technique starts with a model containing a smaller set of extreme points $\bar{\mathcal{P}} \subseteq \mathcal{P}$, the so-called (Restricted) Master Problem (RMP), then it iteratively solves the RMP and an additional Pricing Problem (PP) that checks if the set of points in $\bar{\mathcal{P}}$ is enough to determine the optimal solution (and therefore the procedure stops) or if (at least) one additional variable needs to be added to the RMP. A key component of CG is therefore an exact pricing procedure.

If the Master Problem is convex (i.e., if $f(x)$ and the $g_i(x)$ are convex), it is possible to use the duality theory to present the pricing problem as a cutting plane separation in the dual of the Master Problem. To do so, we can apply Wolfe duality to obtain the following dual of (F-NLM) (see [40] and Appendix B for details about its calculation):

$$
\begin{aligned}
\max \quad & f(x) + \sum_{i \in I} \mu_i g_i(x) + \pi^\top x - \pi_0 & & \\
\text{s.t.} \quad & \nabla_x f(x) + \sum_{i \in I} \mu_i \nabla_x g_i(x) + \pi = 0 & & \\
& -\pi^\top x^p + \pi_0 \geq 0 & \forall p \in \mathcal{P} & \qquad (7) \\
& y^p \geq 0 & \forall p \in \mathcal{P} & \qquad (8) \\
& \mu_i \geq 0 & i \in I & \qquad (9)
\end{aligned}
$$

Starting with a subset of y variables in the primal is equivalent to having only a subset of constraints (7) in the dual. For a given primal-dual solution of the RMP, the pricing problem in the primal is equivalent to a separation problem of constraints (7) in the dual. More formally, let (μ^*, π^*, π_0^*) be an optimal dual solution of the RMP, the pricing problem reduces to find a $p \in \mathcal{P} \setminus \bar{\mathcal{P}}$ such that $-\pi^{*\top} x^p + \pi_0^* < 0$ (or to prove that none exists). This can be done by solving the following model:

$$\begin{aligned} \min \quad & -\pi^{*\top} x + \pi_0^* & \\ \text{s.t.} \quad & h_j(x) \le 0 & j \in J \\ & x \in \{0,1\}^n & \end{aligned}$$

and, if its optimal value is strictly lower than zero, its optimal solution identifies a violated inequality to be added in the dual (or, analogously, a new variable to be added in the primal).

DWR with a Pricing Problem with nonlinear objective function. An alternative reformulation can be obtained if we want to express the objective function and the constraints as linear combination of their value on the extreme points $p \in \mathcal{P}$:

$$\begin{aligned} \text{(F-NLP)} \quad \min \quad & \sum_{p \in \mathcal{P}} f(x^p) y_p & \\ \text{s.t.} \quad & \sum_{p \in \mathcal{P}} g_i(x^p) y_p \le 0 & i \in I \; [\mu_i] \\ & \sum_{p \in \mathcal{P}} y^p = 1 & [\pi_0] \\ & y^p \ge 0 & \forall p \in \mathcal{P} \end{aligned}$$

where μ_i and π_0 denote the dual variables of the associated constraints. In this case the Master Problem is linear, we can therefore use the standard Linear Programming Duality to obtain its dual:

$$\begin{aligned} \max \quad & \pi_0 & \\ \text{s.t.} \quad & \pi_0 - \sum_{i \in I} g_i(x^p) \mu_i \le f(x^p) & \forall p \in \mathcal{P} \\ & \mu_i \ge 0 & i \in I \; . \end{aligned}$$

The pricing problem in this case reduces to find, given a set of optimal dual variables (μ^*, π_0^*), a $p \in \mathcal{P} \setminus \bar{\mathcal{P}}$ such that $\pi_0^* - \sum_{i \in I} g_i(x^p) \mu_i^* > f(x^p)$. This can be done by solving the following model:

$$\begin{aligned} \max \quad & \pi_0^* - \sum_{i \in I} g_i(x) \mu_i^* - f(x) & \\ \text{s.t.} \quad & h_j(x) \le 0 & j \in J \\ & x \in \{0,1\}^n & \end{aligned}$$

that turns out to be a nonlinear problem. Therefore, we refer to the reformulation (F-NLP) as *DWR with a nonlinear Pricing Problem.* We remark that we do not need to impose to $f(x)$ and $g_i(x)$ to be convex to use the duality theory to define the pricing problem as with (F-NLM) because in this case the Master Problem is always linear (and consequently convex).

Reinterpretation of DWR. In order to compare the DWR with nonlinear Master and nonlinear Pricing approaches, we cast both of them in the context of Lagrangian duality. Let us consider the following rewriting of (BF):

$$
\begin{array}{lll}
(\mathrm{BF}') & \min & f(w) \\
& \text{s.t.} & x - w = 0 \qquad\qquad\qquad\qquad [\phi] \qquad\qquad\qquad (10) \\
& & g_i(w) \le 0 \qquad\qquad\qquad\quad i \in I\ [\mu_i] \qquad\qquad (11) \\
& & h_j(x) \le 0 \qquad\qquad\qquad\quad j \in J \\
& & x \in \{0,1\}^n
\end{array}
$$

obtained by decoupling the objective function and the constraints $h_j(x)$ with the introduction of the new set of variables w. Let (F′) be the continuous relaxation of (BF′).

The following theorem shows how the proposed models can be viewed in terms of Lagrangian Duality.

Theorem 1. *When the $f(x)$ and $g_i(x)$ (for all $i \in I$) functions are convex and the Slater condition holds for (F), the bound given by formulation (F-NLM) is equivalent to that given by the Lagrangian Dual of (F′) when constraints* (10) *and* (11) *are relaxed.*

Proof. We prove the equivalence between the Wolfe dual of (F-NLM) and the Lagrangial dual of (F′) for this particular relaxation. In fact, the Lagrangian Dual obtained after relaxing the first two constraints is the following:

$$
\max_{\substack{\mu \in \mathbb{R}_+^{|I|} \\ \phi \in \mathbb{R}^n}} \theta(\mu,\phi) = \max_{\substack{\mu \in \mathbb{R}_+^{|I|} \\ \phi \in \mathbb{R}^n}} \min_{\substack{x^p \in \mathcal{P} \\ w \in \mathbb{R}^n}} L(x^p, w, \mu, \phi) = \max_{\substack{\mu \in \mathbb{R}_+^{|I|} \\ \phi \in \mathbb{R}^n}} \min_{\substack{x^p \in \mathcal{P} \\ w \in \mathbb{R}^n}} \left(f(w) + \phi^\top (x^p - w) + \mu^\top g(w) \right)
$$

that can be rewritten as follows:

$$
\begin{array}{lll}
\max & \left(\sigma + \min\limits_{w \in \mathbb{R}^n} (f(w) - \phi^\top w + \mu^\top g(w)\right) & \\
\text{s.t.} & \sigma - \phi^\top x^p \le 0 & \forall p \in \mathcal{P} \\
& \mu \in \mathbb{R}_+^{|I|}, \phi \in \mathbb{R}^n, \sigma \in \mathbb{R}\ . &
\end{array}
$$

The inner minimization problem is convex unconstrained, therefore it attains the optimum only if its gradient is zero. This allows to rewrite it as the following model:

$$
\begin{array}{lll}
\max & \sigma + f(w) - \phi^\top w + \mu^\top g(w) & \\
\text{s.t.} & \nabla_w f(w) - \phi + \mu^\top \nabla_w g(w) = 0 & \\
& \sigma - \phi^\top x^p \le 0 & \forall p \in \mathcal{P} \\
& \mu \in \mathbb{R}_+^{|I|}, \phi \in \mathbb{R}^n, \sigma \in \mathbb{R}, &
\end{array}
$$

that, after the variable substitutions $\pi_0 = -\sigma$, $\pi = -\phi$, is equal to the Wolfe dual of (F-NLM).

□

Theorem 2. *Formulation (F-NLP) is equivalent to the Lagrangian Dual of (F′) when constraints* (11) *are relaxed.*

Proof. We prove that (F-NLP) turns out to be the dual of the Lagrangian dual of (F′). In fact, we first observe that, if we do not relax constraints (10), (F′) can be remapped to (F) by simple replacement of the w variables. As previously mentioned, in the Lagrangian Dual of (F) where constraints $g_i(x) \leq 0$ are dualized, the set Ω can be replaced by the finite set of points $\mathcal{P}$. In this special case, the Lagrangian Dual can be rewritten as follows:

$$\max_{\mu\in\mathbb{R}_+^{|I|}} \theta(\mu) = \max_{\mu\in\mathbb{R}_+^{|I|}} \min_{\substack{x\in\{0,1\}^n \\ h_j(x)\leq 0, j\in J}} L(x,\mu) = \max_{\mu\in\mathbb{R}_+^{|I|}} \min_{x^p\in\mathcal{P}} L(x^p,\mu) = \max_{\mu\in\mathbb{R}_+^{|I|}} \min_{x^p\in\mathcal{P}} (f(x^p) + \mu^\top g(x^p))$$

which, in turn, can be written as a linear program as:

$$\begin{aligned} \max_{\mu\in\mathbb{R}_+^{|I|}} \theta(\mu) = \max_{\mu\in\mathbb{R}_+^{|I|}} \quad & \sigma \\ \text{s.t.} \quad & \sigma - \mu^\top g(x^p) \leq f(x^p) && \forall p \in \mathcal{P} \end{aligned}$$

and its dual becomes:

$$\begin{aligned} \min \quad & \sum_{x^p\in\mathcal{P}} f(x^p) y^p \\ \text{s.t.} \quad & \sum_{x^p\in\mathcal{P}} g_i(x^p) y^p \leq 0 && \forall i \in I \\ & \sum_{x^p\in\mathcal{P}} y^p = 1 \\ & y^p \geq 0 && \forall p \in \mathcal{P} . \end{aligned}$$

□

The following corollary establishes a relation between DWR with a nonlinear Master Problem and the one with a nonlinear Pricing Problem:

Corollary 1. *When the $f(x)$ and $g_i(x)$ (for all $i \in I$) functions are convex and the Slater condition holds for (F), (F-NLP) is stronger than (F-NLM).*

Proof. The statement is a direct consequence of Theorem 1 and Theorem 2: both formulations can be viewed as Lagrangian Duals of the same formulation (F′). However, (F-NLP) relaxes less constraints than (F-NLM) and therefore it is stronger. □

Corollary 1 states that, when the problem is convex, minimising $f(x)$ on the convex hull of the integer solutions provides a stronger bound than minimising the convex envelope of $f(x)$ on that set. This is not surprising and it has been showed before in other context (See for example Theorem 2.11 in [28]).

4 New reformulations for BQP

In the previous sections we showed how the continuous relaxation of a quadratic problem, reformulated with the QCR method and with DWR, can be seen as particular cases of Langrangian Duals. In principle, these two procedures do not dominate each other because they arise from different relaxations. In this section, we explore the possibility of combining them, applying DWR to an even stronger formulation of (BF), obtained when $f(x)$ is replaced by $f_{\delta,\rho}(x)$.

The discussion presented in Section 3 concerns generic nonlinear problems. Nevertheless, in the following we focus on (BQ), a special case of (BF) where the objective function is quadratic (not necessarily convex) and all constraints are linear:

$$
\begin{aligned}
\text{(BQ)} \quad \min \ & f(x) = x^\top Q x + L^\top x \\
\text{s.t.} \ & Gx \le g \\
& Hx \le h \\
& x \in \{0,1\}^n \ .
\end{aligned}
$$

With a slight abuse of notation, we assume that a subset of constraints $Gx \le g$ and $Hx \le h$ can be equations. When needed, we identify with $G_= x = g_=$ and $G_\le x \le g_\le$ (resp. $H_= x = h_=$ and $H_\le x \le h_\le$) the subset of equations and inequalities of the set of constraints $Gx \le g$ (resp. $Hx \le h$). Moreover, we recall that we identify with $A_= x = b_=$ the union of both sets of equations.

For consistency, we denote by (BQ-QM) the Dantzig-Wolfe Reformulation of (BQ) with nonlinear (more precisely, quadratic) Master Problem, by (BQ-QP) the Dantzig-Wolfe Reformulation of (BQ) with nonlinear (more precisely, quadratic) Pricing Problem, and with (Q-QM) and (Q-QP) the corresponding continuous relaxations.

Finally, we denote by ($BQ_{\delta,\rho}$) the model obtained from (BQ) by replacing $f(x)$ with $f_{\delta,\rho}(x)$ and by ($BQ_{\delta,\rho,\Gamma}$) that obtained by replacing $f(x)$ with $f_{\delta,\rho,\Gamma}(x,z)$. Consequently, we denote by ($Q_{\delta,\rho}$-QM) the DWR of ($BQ_{\delta,\rho}$) with a quadratic Master Problem, and with ($Q_{\delta,\rho}$-QP) the DWR of ($BQ_{\delta,\rho}$) with a quadratic Pricing Problem. A similar notation, with ($Q_{\delta,\rho,\Gamma}$) in place of ($Q_{\delta,\rho}$), is used when $f_{\delta,\rho}(x)$ is replaced by $f_{\delta,\rho,\Gamma}(x,z)$. To complete notation, (Q-QM) and (Q-QP) refer to the DWR of (BQ) with no convexification of the objective $f(x)$.

For completeness, in Table 1 we report an overview of the possible (BQP) relaxations arising from the combination of convexification and strengthening. The Table is organized as follows: DWR options are reported in rows. The different options for the objective function are reported in columns: column $f(x)$ refers to the use of the original (possibily non convex) objective function, while column $f_{\delta,\rho}(x)$ and $f_{\delta,\rho,\Gamma}(x)$ refer to the option of reformulating the objective function with a QCR-like method.

In Appendix C we show a reinterpretation of the reformulations listed in Table 1 in light of Lagrangian Duality theory and based on the results of Section 2 and Section 3.

		Objective function		
		$f(x)$	$f_{\delta,\rho}(x)$	$f_{\delta,\rho,\Gamma}(x,z)$
DWR	QM	(Q-QM)	($Q_{\delta,\rho}$-QM)	($Q_{\delta,\rho,\Gamma}$-QM)
	QP	(Q-QP)	($Q_{\delta,\rho}$-QP)	($Q_{\delta,\rho,\Gamma}$-QP)

Table 1: Summary of possible (BQP) relaxations combining strengthening and convexification for (BQ).

Applying $f_{\delta,\rho}(x)$ to DWR with a quadratic Master Problem. Following the proof of Theorem 1, ($Q_{\delta,\rho}$-QM) is equivalent to the following Lagrangian Dual:

$$\max_{\substack{\delta\in\mathbb{R}^n_+,\rho\in\mathbb{R}\\ \mu\in\mathbb{R}^{|I|}_+\\ \phi\in\mathbb{R}^n}} \min_{\substack{x^p\in\mathcal{P}\\ w\in\mathbb{R}^n}} \left(f_{\delta,\rho}(w) + \phi^\top(x^p - w) + \mu^\top(Gx^p - g)\right) \tag{12}$$

that can be viewed as a special case of the Lagrangian Dual (4) where the set $\{x : Ax \geq b\}$ takes the form:

$$\{(x,y) \in \mathbb{R}^n \times \mathbb{R}^{|\mathcal{P}|}_+ : x - \sum_{p\in\mathcal{P}} x^p y^p = 0, \sum_{p\in\mathcal{P}} y^p = 1\}\ .$$

This implies that, in principle, the optimal dual multipliers can be obtained by solving an additional SDP with an exponential number of variables. This option is clearly prohibitive in practice because it is well-known that SDP solvers do not scale well with the increase of the size of the problem and no effective warm-start techniques exist for SDPs. Nevertheless, we recall that the multipliers (δ^*, ρ^*) (obtained after solving the standard Lagrangian Dual for the QCR method, see Section 2.2) lead to a feasible solution of the Lagrangian Dual (12) and thus can be used to obtain a feasible reformulation of problem (F) of DWR with a nonlinear master problem. For this reason in the following we limit our analysis to reformulations using such multipliers.

Applying $f_{\delta,\rho}(x)$ to DWR with a quadratic Pricing Problem. Also in this case, we investigate the problem of finding the set of multipliers yielding the strongest ($Q_{\delta,\rho}$-QP) relaxation. We first observe the following:

Proposition 1. *For a given $f(x)$ and $\rho' \leq \rho''$, we have that ($Q_{\delta,\rho''}$-QP) is stronger than ($Q_{\delta,\rho'}$-QP).*

Proof. For a given δ and ρ, let $c^p = f_{\delta,\rho}(x^p)$ be the coefficient of a y_p variable (see Section 3). The formula for c^p is the following:

$$c^p = f_{\delta,\rho}(x^p) = f(x^p) + \sum_{j\in J} \delta_j((x^p_j)^2 - x^p_j) + \rho(A_= x^p - b_=)^2\ .$$

The extreme points $x^p \in \mathcal{P}$ take binary coefficients, therefore the quantities $\delta_j((x^p_j)^2 - x^p_j)$ are always equal to zero, regardless of the values of the δ_j. On the other hand, the sign of $\rho(A_= x^p - b_=)^2$ is equal to the sign of ρ, and the value of all the coefficients c^p increases with the increase of ρ proving the claim. □

The following lemma gives an answer to the question of finding the optimal multipliers:

Lemma 1. *Let M be a sufficiently large number, the multipliers δ and ρ providing the strongest dual bound belong to the following set:*

$$\{(\delta,\rho) \in \mathbb{R}^n \times \mathbb{R} : \rho \geq M\}$$

and for such values, ($Q_{\delta,\rho}$-QP) respects the set of equations $A_= x^p = b_=$.

Proof. We can write the problem of finding the best multipliers as follows:

$$\begin{aligned}
\max_{\delta\in\mathbb{R}^n,\rho\in\mathbb{R}} \quad \min \quad & \sum_{p\in\mathcal{P}} c^p y_p \\
\text{s.t.} \quad & c^p = f_{\delta,\rho}(x^p) = f(x^p) + \sum_{j\in J} \delta_j((x_j^p)^2 - x_j^p) + \rho(A_= x^p - b_=)^2 \\
& \sum_{p\in\mathcal{P}} g_i(x^p) y_p \leq 0 && i \in I \\
& \sum_{p\in\mathcal{P}} y^p = 1 \\
& y^p \geq 0 && \forall p \in \mathcal{P} .
\end{aligned}$$

As showed in the proof of Proposition 1, the value of all the coefficients c^p increases with the increase of ρ. If ρ is sufficiently large, all the extreme points with the quantity $A_= x^p - b_=$ different from zero have a corresponding cost $c^p = +\infty$ and they are never part of an optimal solution. For this reason, starting from a value of $\rho \geq M$, only the variables y^p corresponding to extreme points that respect also the constraints $A_= x^p = b_=$ are part of the optimal solution of ($Q_{\delta,\rho}$-QP). □

The following theorems show how ($Q_{\delta,\rho}$-QP) behaves in the extreme cases where the Master Problem contains either only or no equations:

Theorem 3. *If $Gx \leq g$ contains only equations, and $\rho \geq M$ (for a sufficiently large value of M), solving ($Q_{\delta,\rho}$-QP) is equivalent to solve the original problem (BQ).*

Proof. The proof directly follows from Lemma 1, observing that no more constraints need to be satisfied. □

Theorem 4. *If $Gx \leq g$ contains only inequalities, solving ($Q_{\delta,\rho}$-QP) is equivalent to solve (Q-QP).*

Proof. We can even prove that, in such a case, ($Q_{\delta,\rho}$-QP) directly reduces to (Q-QP). The proof is similar to the one of Lemma 1: if no equations are present

$$c^p = f_{\delta,\rho}(x^p) = f(x^p) + \sum_{j\in J} \delta_j((x_j^p)^2 - x_j^p)$$

but since x^p have binary coefficients, $c^p = f(x^p)$. □

Lemma 1 implies that the multipliers giving the strongest reformulation are easy to characterize. However, in Section 6.3 we show that they are not useful in practice.

Finally we notice that, while Corollary 1 implies that ($Q_{\delta,\rho}$-QP) is stronger than ($Q_{\delta,\rho}$-QM) for any suitable choice of δ and ρ, such a theoretical guarantee applies to (Q-QP) and (Q-QM) only when the problem is convex. In fact, in the general case we prove the following.

Proposition 2. *There is no dominance between (Q-QP) and (Q-QM).*

Proof. Let us consider the two following instances:

$$
\begin{array}{lll}
(\mathcal{A})\min & f(x) = -x_1 \cdot x_2 \\
\text{s.t.} & g(x) : x_1 + x_2 = 1 \\
& h(x) : x_1 + x_2 \le \mathbf{1} \\
& x_1, x_2 \in \{0,1\}
\end{array}
\qquad
\begin{array}{lll}
(\mathcal{B})\min & f(x) = -x_1 \cdot x_2 \\
\text{s.t.} & g(x) : x_1 + x_2 = 1 \\
& h(x) : x_1 + x_2 \le \mathbf{2} \\
& x_1, x_2 \in \{0,1\}
\end{array}
$$

We recall that $h(x)$ are the constraints that we want to strengthen. Let $\Omega_{\mathcal{A}}$ be the convex hull of $\{x : x_1 + x_2 \le 1, x \in \{0,1\}^2\}$ and $\Omega_{\mathcal{B}}$ be the convex hull of $\{x : x_1 + x_2 \le 2, x \in \{0,1\}^2\}$. Let $\mathcal{P}_{\mathcal{A}}$ and $\mathcal{P}_{\mathcal{B}}$ be the sets of extreme points of $\Omega_{\mathcal{A}}$ and $\Omega_{\mathcal{B}}$. We have that :

$$\mathcal{P}_{\mathcal{A}} = \{(0,0),(1,0),(0,1)\}, \mathcal{P}_{\mathcal{B}} = \{(0,0),(1,0),(0,1),(\mathbf{1},\mathbf{1})\} .$$

The optimal integer solution is 0 for both ($\mathcal{A}$) and ($\mathcal{B}$). The optimal solution for both $(\mathcal{A}-QM)$ and $(\mathcal{B}-QM)$ is attained when $y_{(1,0)} = y_{(0,1)} = x_1 = x_2 = 0.5$, and has a value of $-0.5 \cdot 0.5 = -0.25$. The optimal solution for $(\mathcal{A}-QP)$ is 0, since all the feasible columns have objective function value 0. On the other hand, in $(\mathcal{B}-QP)$ column $(1,1)$ has objective function value -1, and therefore the optimal solution is attained when $y_{(0,0)} = y_{(1,1)} = 0.5$, and has a value of $0 \cdot 0.5 - 1 \cdot 0.5 = -0.5$. Therefore, $(\mathcal{A}-QP)$ is stronger than $(\mathcal{A}-QM)$, while $(\mathcal{B}-QM)$ is stronger than $(\mathcal{B}-QP)$. □

Proposition 3. *There is no dominance between (Q-QP) and ($Q_{\delta,\rho}$-QM).*

Proof. Let us consider the same pair of instances ($\mathcal{A}$) and ($\mathcal{B}$) introduced in the proof of Proposition 1. Since $(\mathcal{A}-QM)$ is stronger than $(\mathcal{A}-QP)$ and $(\mathcal{A}_{\delta,\rho}-QM)$ is stronger than $(\mathcal{A}-QM)$, it also holds that $(\mathcal{A}_{\delta,\rho}-QM)$ is stronger than $(\mathcal{A}-QP)$. As an example of the converse behaviour, let us consider the following:

$$
\begin{array}{ll}
(\mathcal{C})\min & f(x) = -x_1 \cdot x_2 - x_3 \cdot x_4 \\
\text{s.t.} & g(x) : x_1 + x_2 + x_3 + x_4 = 1 \\
& h(x) : 2 \cdot x_1 + x_2 + x_3 + 2 \cdot x_4 \le 2 \\
& x_1, x_2, x_3, x_4 \in \{0,1\}
\end{array}
$$

Let $\Omega_{\mathcal{C}}$ be the convex hull of $\{x : 2 \cdot x_1 + x_2 + x_3 + 2 \cdot x_4 \le 2, x \in \{0,1\}^4\}$ and $\mathcal{P}_{\mathcal{C}}$ be the set of extreme points of $\Omega_{\mathcal{C}}$. We have that:

$$\mathcal{P}_C = \{(0,0,0,0),(1,0,0,0),(0,1,0,0),(0,0,1,0),(0,0,0,1),(0,1,1,0)\}.$$

The optimal integer solution is 0. The optimal solution for $(C-QP)$ is 0, since all the feasible columns have objective value 0. On the other hand, the QCR method yields the following set of optimal multipliers: $\delta^* = \{0.5, 0.5, 0.5, 0.5\}$ (due to the cardinality constraint, the value of ρ^* has no influence). The objective function of $(C_{\delta,\rho}-QM)$ becomes

$$f_{\delta,\rho}(x) = -x_1 \cdot x_2 - x_3 \cdot x_4 + 0.5(x_1^2 - x_1) + 0.5(x_2^2 - x_2) + 0.5(x_3^2 - x_3) + 0.5(x_4^2 - x_4).$$

In this case, an optimal solution for $(C_{\delta,\rho}-QM)$ is attained when $y_{(1,0,0,0)} = y_{(0,1,0,0)} = x_1 = x_2 = 0.5$, with value -0.5. Therefore $(C-QP)$ is stronger than $(C_{\delta,\rho}-QM)$. $\square$

Using $f_{\delta,\rho,\Gamma}(x,z)$ instead of $f_{\delta,\rho}(x)$. The following theorem shows the effect of improving the convexification of the objective function on DWR with a quadratic Pricing Problem:

Theorem 5. *For a generic $f(x)$, and for any set of multipliers (δ,ρ), ($Q_{\delta,\rho}$-QP) and ($Q_{\delta,\rho,\Gamma}$-QP) are equivalent.*

Proof. Both formulations differ only in terms of their objective functions. The coefficients $\hat{c}^p$ of the y_p variables for ($Q_{\delta,\rho,\Gamma}$-QP) are the following:

$$\hat{c}^p = f_{\delta,\rho,\Gamma}(x^p, z^p) = f(x^p) + \sum_{j\in J} \delta_j((x_j^p)^2 - x_j^p) + \rho(A_= x^p - b_=)^2 + \sum_{i\in J}\sum_{j\in J} \Gamma_{ij}(z_{ij}^p - x_i^p x_j^p) \ .$$

Since the extreme points x^p are always binary, we have $z_{ij} = x_i x_j$, therefore $\sum_{i\in J}\sum_{j\in J} \Gamma_{ij}(z_{ij}^p - x_i^p x_j^p) = 0$. This implies that $\hat{c}^p = f(x^p) + \sum_{j\in J} \delta_j((x_j^p)^2 - x_j^p) + \rho(A_= x^p - b_=)^2$, that corresponds to the coefficients c^p of the y_p variables for ($Q_{\delta,\rho}$-QP). Showing that the two formulations are equivalent. $\square$

Theorem 5 allows us to exclude the family of formulations ($Q_{\delta,\rho,\Gamma}$-QP) from our computational analysis, being equivalent to ($Q_{\delta,\rho}$-QP) but larger in terms of number of variables and constraints. We remark that, for a similar reasoning, any value of δ provides the same values for the coefficients $\hat{c}^p$. However, we keep δ in the notation for coherence with the other QCR formulations.

Overview of the reformulations. It is possible to establish a hierarchy among the reformulations, based on their strength.

The overall hierarchy is provided in Figure 1: for each pair of formulation families A and B, we depict $A \Rightarrow B$ when A is stronger than B, $A \Leftrightarrow B$ when A and B are equally strong, and $A \not\Leftrightarrow B$ when there is no dominance between A and B. Near each dominance arrow we provide the proposition proving the statement.

By looking at Figure 1 we see that ($Q_{\delta,\rho}$-QP) and ($Q_{\delta,\rho,\Gamma}$-QP) are the ones providing the strongest bounds.

$$
\begin{array}{ccccc}
(\text{Q-QM}) & \Leftarrow_{[8],\delta^*,\rho^*} & (\text{Q}_{\delta,\rho}\text{-QM}) & \Leftarrow_{[7],\delta^*,\rho^*,\Gamma^*} & (\text{Q}_{\delta,\rho,\Gamma}\text{-QM}) \\
\not\Updownarrow_{\text{Prop.2}} & \not\nearrow\ \text{Prop.3} & \Uparrow_{\text{Cor.1}} & & \Uparrow_{\text{Cor.1}} \\
(\text{Q-QP}) & \Leftarrow_{\text{Prop.1},\rho>0} & (\text{Q}_{\delta,\rho}\text{-QP}) & \Leftrightarrow_{\text{Th.5}} & (\text{Q}_{\delta,\rho,\Gamma}\text{-QP})
\end{array}
$$

Figure 1: Hierarchy of reformulations.

5 Case study: the 0-1 exact k-item Quadratic Knapsack Problem (kQKP)

As a case study we focus on the 0-1 exact k-item Quadratic Knapsack Problem [29, 30] (kQKP). The kQKP is a BQP with a generic objective function, but only two constraints (one equation and one inequality). It is hence perfect to be used as test problem, allowing us to focus on the methodology rather than the technical issues.
More precisely, the kQKP consists of maximizing a quadratic function subject to a capacity and a cardinality constraint, which are both linear. It can be formulated as follows:

$$
\begin{array}{llll}
(BK) & \max & f(x) = \displaystyle\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}x_ix_j & \\
& \text{s.t.} & \displaystyle\sum_{j=1}^{n} a_jx_j \leq b & (kn) \\
& & \displaystyle\sum_{j=1}^{n} x_j = k & (ca) \\
& & x_j \in \{0,1\} & j = 1,\dots,n
\end{array}
$$

where n (number of items), k (number of items to be filled in the knapsack), a_j (weight of item j), c_{ij} (profit associated with the selection of items i and j) and b (capacity of the knapsack) are nonnegative integers. Without loss of generality, the matrix $C = (c_{ij})$ is assumed to be symmetric.
Moreover, we assume that $\max_{j=1,\dots,n} a_j \leq b < \sum_{j=1}^{n} a_j$ in order to avoid either trivial solutions or trivial variable fixing via constraint (kn). Let us denote by k_{max} the largest number of items which could be filled in the knapsack, that is the largest k such that the sum of the k smallest a_j values does not exceed b. Therefore, we can assume that $k \in \{2,\dots,k_{max}\}$, where k_{max} can be found in O(n) time [4, 15]. Outside this range, either the value of the problem is equal to $\max_{i=1,\dots,n} c_{ii}$ (for k=1), or the domain of (BK) is empty (for $k > k_{max}$). For notation consistency, we refer to (K) as the continuous relaxation counterpart of formulation (BK).

5.1 Formulations Overview

In Table 2 we present an overview of the possible options arising from the combination of convexification and strengthening for (BK). The Table is organized as Table 1: DWR options are reported in rows, with QM and QP refering to the formulations with either a quadratic Master Problem or a quadratic Pricing Problem. Moreover, (kn),

		constraints in the pricing	$f(x)$	Objective function $f_{\delta,\rho}(x)$	$f_{\delta,\rho,\Gamma}(x,z)$
DWR	QM	(kn)	`NCM`	($\mathrm{K}_{\delta,\rho}$-QM-kn)	($\mathrm{K}_{\delta,\rho,\Gamma}$-QM-$kn$)
		(ca)	`NCM`	`IPP`	`IPP`
		(kn,ca)	`NCM`	($\mathrm{K}_{\delta,\rho}$-QM-kn,ca)	($\mathrm{K}_{\delta,\rho,\Gamma}$-QM-$kn,ca$)
	QP	(kn)	(K-QP-kn)	($\mathrm{K}_{\delta,\rho}$-QP-kn)	($\mathrm{K}_{\delta,\rho,\Gamma}$-QP-$kn$)
		(ca)	(K-QP-ca)	($\mathrm{K}_{\delta,\rho}$-QP-ca)	($\mathrm{K}_{\delta,\rho,\Gamma}$-QP-$ca$)
		(kn,ca)	`POP`	`POP`	`POP`

Table 2: Summary of possible formulations combining strengthening and convexification for (BK).

(ca) and (kn,ca) refer to the possible choices of constraints to be strengthened, that is the choice of Ω. Also in this case, the different options for the objective function are reported in columns: column $f(x)$ refers to the use of the original (possibily non convex) objective function, while column $f_{\delta,\rho}(x)$ and $f_{\delta,\rho,\Gamma}(x)$ refer to the option of reformulating the objective function with a QCR-like method.

We first note that not every combination is relevant:

- `NCM`
  The combination marked with `NCM` (NonConvex Master) are those leading to a possibly nonconvex Master Problem. On these cases it is not possible, in general, to ensure through KKT conditions that the solution achieved during master optimization steps is a global optimum, and therefore it is not possible to guarantee global convergence of the column generation process. For this reason we discard these combinations from our analysis.

- `IPP`
  The combination marked with `IPP` (Integrality Property on Pricing) are those where the Pricing Problem has the so-called *integrality property*, i.e. the continuous relaxation of the Pricing Problem has always an optimal integer solution. In fact, the Pricing Problem is linear, and amounts in selecting a set of k items of maximum prize. That is, even if we apply DWR, no improvement in the continuous relaxation bound is obtained. We therefore decide to discard these configurations.

- `POP`
  The combinations marked with `POP` (Pricing is Original Problem) are those where the Pricing Problem presents a quadratic objective function, and the same feasible region as the original problem. Therefore, a single column generation iteration would be as hard as the original problem. Also these configurations are discarded from further analysis.

Additionally, a preliminary set of experiments among ($\mathrm{K}_{\delta,\rho}$-QP-ca) and ($\mathrm{K}_{\delta,\rho}$-QP-kn), showed the latter to produce more interesting computational insights.

Therefore, among the remaining combinations, we show in details the master and pricing problems of ($\mathrm{K}_{\delta,\rho,\Gamma}$-QM-$kn,ca$) and ($\mathrm{K}_{\delta,\rho}$-QP-$kn$), being representative of the whole set of options.

Indeed, it is worth noting that ($\mathrm{K}_{\delta,\rho,\Gamma}$-QM-$kn,ca$) provides the strongest reformulation of the feasible region, while the Pricing Problem is kept of manageable complexity by leaving the quadratic part of the objective function in the master.

$$
\begin{array}{lllll}
(\mathrm{K}_{\delta,\rho,\Gamma}\text{-QM-}kn,ca) & \max & f_{\delta,\rho,\Gamma}(x,z) & & \\
& \text{s.t.} & x_j = \sum\limits_{p\in\mathcal{P}_{kn,ca}} x_j^p y^p & j = 1,\ldots,n & [\pi_j] \\
& & z_{ij} = \sum\limits_{p\in\mathcal{P}_{kn,ca}} z_{ij}^p y^p & i,j = 1,\ldots,n & [\xi_{ij}] \\
& & \sum\limits_{p\in\mathcal{P}_{kn,ca}} y^p = 1 & & [\pi_0] \\
& & y^p \geq 0 & p \in \mathcal{P}_{kn,ca} &
\end{array}
$$

with $\mathcal{P}_{kn,ca}$ being the set of extreme points of

$$
\begin{aligned}
\mathrm{conv}\{(x,z) | & \sum_{j=1}^n a_j x_j \leq b; \sum_{j=1}^n x_j = k; \\
& z_{ij} \leq x_i,\ z_{ij} \leq x_j, z_{ij} \geq 0,\ z_{ij} \geq x_i + x_j - 1, \qquad i,\ j = 1,\ldots,n \\
& x \in \{0,1\}^n\}
\end{aligned}
$$

and π, ξ, π_0 being the dual variables associated to the constraints in the master. We observe that the bounds on the x and z variables are automatically respected by the variables y^p associated to the extreme points of $\mathcal{P}_{kn,ca}$. Let π^*, ξ^*, π_0^* be the optimal dual variables of a master solution during a generic column generation iteration, the Pricing Problem can be written as follows:

$$
\begin{array}{lll}
\max & \sum\limits_{j=1}^n \pi_j^* x_j + \sum\limits_{i,j=1}^n \xi_{ij}^* z_{ij} + \pi_0^* & \\
\text{s.t.} & \sum\limits_{j=1}^n a_j x_j \leq b & \\
& \sum\limits_{j=1}^n x_j = k & \\
& z_{ij} \leq x_i,\ z_{ij} \leq x_j & i,j = 1,\ldots,n \\
& z_{ij} \geq 0,\ z_{ij} \geq x_i + x_j - 1 & i,j = 1,\ldots,n \\
& x_j \in \{0,1\} & j = 1,\ldots,n.
\end{array}
$$

On the other hand, in ($\mathrm{K}_{\delta,\rho}$-QP-kn) the Master Problem is linear and contains only two constraints, and the Pricing Problem is a binary Quadratic Knapsack Problem. This makes such a case potentially appealing to be used in dedicated algorithms, since effective ad-hoc combinatorial algorithms exist for pricing [9, 11, 33, 34]:

$$
\begin{array}{llll}
(\mathrm{K}_{\delta,\rho}\text{-QP-}kn) & \max & \sum_{p\in\mathcal{P}_{\mathrm{kn}}} c^p_{\delta^*,\rho^*} y^p & \\
& \text{s.t.} & \sum_{p\in\mathcal{P}_{\mathrm{kn}}} \sum_{j=1}^{n} x^p_j y^p = k & [\mu] \\
& & \sum_{p\in\mathcal{P}_{\mathrm{kn}}} y^p = 1 & [\pi_0] \\
& & y^p \geq 0 \qquad p \in \mathcal{P}_{\mathrm{kn}} &
\end{array}
$$

with $\mathcal{P}_{\mathrm{kn}}$ being the set of extreme points of $conv\{x | \sum_{j=1}^{n} a_j x_j \leq b, x \in \{0,1\}^n\}$, and μ and π_0 being the dual variables associated to the constraints in the Master Problem.

Denoting as μ^* and π_0^* the optimal dual variables of a master solution during a column generation iteration, the Pricing Problem can be written as follows:

$$
\begin{array}{lll}
\max & f_{\delta^*,\rho^*}(x) + \sum_{i=1}^{n} \mu_i^* x_i + \pi_0^* & \\
\text{s.t.} & \sum_{j=1}^{n} a_j x_j \leq b & \\
& x_j \in \{0,1\} & j = 1, \ldots, n \, .
\end{array}
$$

6 Computational comparison

Building on the results of the previous sections, we designed an experimental campaign in order to investigate the computational properties of our reformulations. Our main aim is to assess the tradeoff that can be obtained in terms of dual bound quality versus computing time.

In this section, the following set of instances from the literature is used as testbed:

- Density-Dependent (`DD`) instances.
  This set consists of kQKP instances produced by the random generator of [29]. The number of items takes the following values: $n = 50, 60, \ldots, 100$; b, a_j and c_{ij} are nonnegative integers. The values of $k \in \{2, \ldots, \min\{k_{max}, \lfloor \frac{n}{4} \rfloor\}\}$ are chosen in a way to make the instances harder than the corresponding Quadratic Knapsack instance. The density of the objective function `Dens` (defined as the percentage ratio between the number of non zero coefficients c_{ij} and n^2) takes the following values: `Dens` $= 25, 50, 75, 100$. For any combination of n and `Dens`, 10 feasible random instances are included in the data set. This test-bed consists in total of 240 instances.

We therefore implemented a `C++` framework, in which the Master Problem is solved by the continuous solver of IBM `CPLEX` [23]. Effective ad-hoc algorithms exist for solving some of the Pricing Problems arising from our reformulations. In fact, in preliminary investigations [12], we experimented on using them. However, to further improve the fairness of our comparisons, we decided not to use them, but rather to always solve Pricing Problems by the integer (either quadratic or linear) solver of IBM

CPLEX. In both cases we kept default settings, except for the MIP emphasis during pricing, which was set to "Hidden Feasibility" for up to 60 seconds, and then set to "Best Bound", returning the first solution found having positive reduced cost. That is, once again the kQKP is used as benchmark problem, but its specific combinatorial structure is not exploited by our optimization algorithms. Such a generic approach is not without drawbacks: convexification methods produce large coefficient values (often in the order of magnitude of 10^9 or more), which in turn affect column generation numerical stability. Special care is required especially at the tail of column generation, when by contrast, very low values of reduced cost are expected as optimal pricing solutions. Hence, fine tuning of numerical constants turned out to be important: choosing a relative optimality tolerance which is too low (resp. too high) may produce a stall in the LP solver (resp. sub-optimal columns to be returned).

In our implementation, when quadratic masters are optimized by the barrier algorithm, the convergence tolerance threshold was initially set to 10^{-9}, and lowered to 10^{-12} when no new column of positive reduced cost was found; setting the duals of convexity contraints as objective offset in pricing also helped to improve stability. Unfortunately, we argue that suitable values for these thresholds are machine (and maybe even operating system) dependent. In fact, we experimented with a few different versions of CPLEX, including 12.8, but release 12.6.3 produced best performances overall.

Our tests run on a PC equipped with an Intel i7 6700K 4.0GHz CPU (4 cores, up to 8 hardware threads), 32GB of RAM, running Ubuntu Linux.

6.1 Dual Bounds overview

In Table 3 the results concerning selected families of formulations are presented. The table is divided into five vertical blocks. The second block, which are ($\mathrm{K}_{\delta^*,\rho^*}$) and ($\mathrm{K}_{\delta^*,\rho^*,\Gamma^*}$), reports the results concerning the models obtained after convexifying the objective function with the techniques introduced in Section 2, without DWR.

As explained in Section 3, to obtain a reformulation amenable to be solved by DWR, the objective function in the Master Problem needs to be convex. For this reason we need to use either $f_{\delta^*,\rho^*}(x)$ or $f_{\delta^*,\rho^*,\Gamma^*}(x,z)$ when dealing with Quadratic Master Problems: the third and forth block, which are ($\mathrm{K}_{\delta^*,\rho^*}$-QM-*kn,ca*) and ($\mathrm{K}_{\delta^*,\rho^*,\Gamma^*}$-QM-*kn,ca*), report the results concerning the models obtained after applying also DWR as explained in Section 4. The results on the two remaining options of DWR with a quadratic Master Problem (i.e., ($\mathrm{K}_{\delta^*,\rho^*,\Gamma^*}$-QM-*kn*) and ($\mathrm{K}_{\delta^*,\rho^*}$-QM-*kn*)) provide no further insights:

Experimental Observation 1. *No significant experimental difference is observed between ($K_{\delta^*,\rho^*,\Gamma^*}$-QM-kn) (resp. ($K_{\delta^*,\rho^*}$-QM-kn)) and ($K_{\delta^*,\rho^*,\Gamma^*}$-QM-kn,ca) (resp. ($K_{\delta^*,\rho^*}$-QM-kn,ca)) neither in terms of duality gap nor in terms of computing time.*

These cases are therefore dropped from the Table.

In the case of Quadratic Pricing Problems, it is not mandatory to convexify the objective function. However, the option of keeping a potentially non-convex objective function is interesting only if ad-hoc algorithms are used for pricing, as otherwise generic convexification techniques are employed by the general purpose MIP solver.

For this reason we include in the table only ($\mathrm{K}_{\delta^*,\rho^*}$-QP-*kn*). We remind to Section 6.3 for an analysis of (K-QP-*kn*) as a special case of ($\mathrm{K}_{\delta^*,\rho^*}$-QP-*kn*) where multipliers take value 0, and to [12] for a computational assessment of using ad-hoc procedures for solving QPs which are potentially non-convex.

We remark that ($\mathrm{K}_{\delta^*,\rho^*}$) and ($\mathrm{K}_{\delta^*,\rho^*,\Gamma^*}$) can be solved in polynomial time, while the remaining models in the table require to optimize a NP-hard subproblem, so they are not directly comparable in terms of scalability. Nevertheless, we include them as reference for comparing the dual bounds obtained.

The table shows horizontally the average of tests concerning the ten instances with the same value of n and `Dens`. For each block using DWR we report the computing (clock) time; relative results do not differ significantly when CPU time is considered instead of clock time. We omit the computing time when DWR is not used: ($\mathrm{K}_{\delta^*,\rho^*}$) appeared to be one order of magnitude faster than ($\mathrm{K}_{\delta^*,\rho^*,\Gamma^*}$), but each test took less than 0.1s. We do not report the pre-processing time needed to compute optimal multipliers because it is neglectable. For all blocks we also report the duality gap (defined as the difference between the continuous relaxation bound and the optimal value, divided by the optimal value). In the last line the total average is reported. An outlook at the trend of computing times yields the following:

Experimental Observation 2. *For all methods, computing times increase significantly as the size of the instance increases, and change only marginally with their density.*

Comparing the two columns of block 2, and blocks 3 with block 4, immediately leads to the following observation:

Experimental Observation 3. *The use of $f_{\delta^*,\rho^*,\Gamma^*}(x,z)$ instead of $f_{\delta^*,\rho^*}(x)$ improves consistently the strength of the dual bound.*

We remark that such a behavior is no longer present in the case of quadratic pricing, for which Theorem 5 states the equivalence of both convexifications.

6.2 DWR with a quadratic Master Problem

As theoretically expected, the duality gaps reported in the third and fifth block (resp. fourth block) are smaller than those in the first column (resp. second column) of the first block, at the expense of higher computing time.

The following observation shows that DWR can be an appealing alternative to complex convexification methods:

Experimental Observation 4. *($K_{\delta^*,\rho^*,\Gamma^*}$) and ($K_{\delta^*,\rho^*}$-QM-kn,ca) yield a similar effect in terms of dual bound improvement.*

The first column (resp. second column) of block one and block three (resp. block four) correspond to formulations which differ in terms of strengthening of the feasible region. Nevertheless, the bound improvement is not as outstanding as expected, and the computing time grows by one order of magnitude. The same applies between blocks three and four, in which a more accurate convexification is performed. This

Dens	n	$(\mathrm{K}\delta^*,\rho^*)$ Gap	$(\mathrm{K}\delta^*,\rho^*,\Gamma^*)$ Gap	$(\mathrm{K}\delta^*,\rho^*$-QM-$kn,ca)$ Time	Gap	$(\mathrm{K}\delta^*,\rho^*,\Gamma^*$-QM-$kn,ca)$ Time	Gap	$(\mathrm{K}\delta^*,\rho^*$-QP-$kn)$ Time	Gap
25%	50	38.5%	30.9%	1.1	38.4%	3.5	29.1%	0.3	0.0%
	60	84.5%	72.0%	1.7	84.4%	3.5	69.2%	0.4	0.0%
	70	42.4%	34.4%	2.5	42.2%	7.7	32.9%	1.0	0.0%
	80	37.9%	30.7%	3.1	37.7%	15.1	29.6%	7.3	0.0%
	90	76.7%	63.8%	3.1	76.7%	14.0	61.2%	15.2	0.0%
	100	60.6%	53.1%	4.1	60.6%	24.9	52.3%	49.9	0.0%
50%	50	31.4%	25.3%	1.4	31.1%	3.3	23.7%	0.3	0.0%
	60	19.8%	15.1%	1.8	19.5%	6.3	14.2%	0.5	0.0%
	70	57.0%	50.9%	2.6	56.9%	11.4	48.9%	3.2	0.0%
	80	64.1%	55.4%	3.0	64.1%	10.5	53.7%	9.6	0.0%
	90	63.8%	57.0%	3.2	63.8%	22.0	55.9%	36.3	0.0%
	100	18.5%	14.2%	4.5	18.5%	29.7	14.0%	107.4	0.0%
75%	50	70.9%	64.6%	1.5	70.9%	2.9	61.9%	0.2	0.0%
	60	86.3%	79.8%	1.5	86.0%	4.1	76.8%	1.0	0.0%
	70	58.2%	50.9%	2.1	58.2%	6.6	49.3%	4.8	0.0%
	80	54.0%	48.7%	2.4	54.0%	14.8	47.7%	39.2	0.0%
	90	42.1%	37.9%	3.6	42.1%	26.4	37.1%	353.5	0.0%
	100	27.2%	21.7%	4.6	27.2%	35.1	21.5%	353.6	0.0%
100%	50	50.4%	45.8%	1.4	50.3%	3.5	44.0%	0.8	0.0%
	60	48.8%	43.6%	1.9	48.8%	5.8	42.4%	1.0	0.0%
	70	44.5%	39.6%	2.5	44.4%	11.7	38.7%	6.9	0.0%
	80	49.2%	43.3%	2.7	49.2%	12.1	42.1%	66.4	0.0%
	90	27.0%	23.9%	3.7	27.0%	24.3	23.4%	69.6	0.0%
	100	59.8%	55.1%	3.5	59.8%	24.3	54.1%	319.8	0.0%
Avg.		50.6%	44.1%	2.6	50.5%	13.5	42.6%	60.3	0.0%

Table 3: DD instances, continuous relaxations of QCR and DWR with quadratic Master and quadratic Pricing.

suggests that while using DWR strengthening, employing sophisticated convexification mechanisms may not be worthwhile.

Among all our tests, the most surprising result is arising from the block related to $(\mathrm{K}_{\delta^*,\rho^*,\Gamma^*}$-QM-$kn,ca)$:

Experimental Observation 5. *The duality gap obtained after applying ($K_{\delta^*,\rho^*,\Gamma^*}$-QM-kn,ca), compared to ($K_{\delta^*,\rho^*,\Gamma^*}$), decreases only from* 44% *to* 42.6%*.*

It is well known in combinatorial optimization that strenghtening a feasible region usually helps significantly in improving the dual bound. For this reason, the modest improvement of the duality gap obtained after strenghtening the knapsack and the cardinality constraints (i.e. optimizing over the convex hull of the whole feasible region) is, at first glance, counter-intuitive. In fact, if the objective function was linear, such a reformulation would always fully close the duality gap.

A possible reason for this behaviour would be that the optimal solution is attained in the interior of the convex hull of the feasible region. In this case, strengthening with DWR yields no improvement. However, in a preliminary experiment, we found that less than 10% of the instances have an optimal solution in the interior of the feasible region.

An alternative explanation of the behaviour showed in Table 3 is related to the fact that $f_{\delta^*,\rho^*}(x)$ and $f_{\delta^*,\rho^*,\Gamma^*}(x,z)$ are only "artificially" convex functions. Such functions are known to be "flatter" than a "real" convex function $f(x)$ (see for example [3]). Intuitively, an objective function is flat if it has descent directions, along which the value of the objective function does not change significantly. For such a function, its value does not change much on the feasible region, making it less interesting to strengthen it with DWR.

6.3 DWR with a quadratic Pricing Problem

The results concerning ($\mathrm{K}_{\delta^*,\rho^*}$-QP-$kn$) are even more insightful.

Experimental Observation 6. *(K_{δ^*,ρ^*}-QP-kn) closes* 100% *of the open gap of all the instances in a reasonable amount of time.*

This last striking result is coherent with the theoretical analysis showed in Section 4, more precisely, with Proposition 1 and Theorem 3.

The design of an exact approach based on the relaxation proposed goes beyond the scope of this work. However, the fact that with ($\mathrm{K}_{\delta^*,\rho^*}$-QP-$kn$) we are always able to close the whole integrality gap suggests that a comparison with the overall solving time of `CPLEX` is pertinent. In Appendix D we undertake such a comparison.

In an effort for understanding the extreme behaviour of ($\mathrm{K}_{\delta^*,\rho^*}$-QP-$kn$), we focus on the objective function of the pricing problem. We recall that, according to Theorem 2 ($\mathrm{K}_{\delta^*,\rho^*}$-QP-$kn$) is in general a relaxation of the kQKP, while according to Theorem 3, for a sufficiently large value of ρ, the pricing problem solves *de facto* a kQKP. So an experimenal question arises, on the impact of choosing pertinent values for ρ. Therefore, in a preliminary experiment we tested the following reformulation of (BK):

$$
\begin{aligned}
(\mathrm{BK}_{\rho^M}) \quad \max \quad & f(x) = \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}x_ix_j - \rho^M \left(\sum_{j=1}^{n} x_j - k\right)^2 \\
\text{s.t.} \quad & \sum_{j=1}^{n} a_j x_j \le b \\
& x \in \{0,1\}^n
\end{aligned}
$$

where ρ^M is a sufficiently large value (that we identified empirically). We do not report the results in detail because, as expected, using (BK_{ρ^M}) instead of (BK) turns out to be, in terms of computing time, way worse.

Summarizing, such a big-M approach proves not to be viable, but at the same time providing sufficiently high penalties for violating the cardinality constraint seems to be a key ingredient for a tight relaxation. We therefore argue that a critical issue is actually the choice of the multiplier ρ, that is by far non trivial. In the following, we report a set of experiments aimed at obtaining a better understanding of such a phenomenon.

Additional experiments on ρ. From a geometric point of view, the multiplier ρ fully controls the convexity of the pricing objective function, as the dual variables act only on its linear part. By setting $\rho = 0$, the pricing objective is convex in none of the `DD` instances, while large values of ρ ensure it to be convex. Indeed, as explained in Appendix A, setting $\rho = \rho^*$ is always sufficient (but not necessary) to make each

Dens	n	Avg. Dual. Gap $\rho = 0$	$\rho = \bar{\rho}$	$\rho = \rho*$	Avg. Time (s) $\rho = 0$	$\rho = \bar{\rho}$	$\rho = \rho*$	Avg. Feasible Cols $\rho = 0$	$\rho = \bar{\rho}$	$\rho = \rho*$	Avg. CG iter. $\rho = 0$	$\rho = \bar{\rho}$	$\rho = \rho*$
25%	50	19.43%	0.00%	0.00%	12.5	1.1	0.3	11.00%	46.14%	53.58%	6.1	7.6	5.2
	60	35.08%	2.60%	0.00%	18.9	0.4	0.4	7.08%	34.78%	49.70%	6.3	7.2	6.1
	70	14.85%	3.67%	0.00%	54.6	10.4	1.1	10.54%	41.56%	50.55%	7.0	6.5	5.8
	80	9.42%	0.00%	0.00%	45.7	6.0	7.3	9.33%	41.41%	56.64%	4.8	7.2	6.6
	90	26.81%	0.00%	0.00%	83.6	32.8	15.2	5.58%	43.19%	59.55%	6.4	9.1	5.6
	100	21.79%	0.00%	0.00%	82.6	68.7	49.9	0.91%	35.30%	62.26%	4.5	10.1	6.7
50%	50	28.89%	0.00%	0.00%	39.4	2.0	0.3	11.06%	54.20%	66.85%	5.0	7.7	5.4
	60	12.77%	0.00%	0.00%	47.1	0.5	0.5	22.86%	62.62%	56.83%	6.2	5.9	5.6
	70	31.91%	0.00%	0.00%	47.9	13.3	3.2	4.44%	45.42%	58.17%	4.7	8.7	6.2
	80	40.87%	2.79%	0.00%	71.9	142.5	9.6	0.00%	32.44%	64.96%	4.6	9.0	6.6
	90	42.91%	0.00%	0.00%	57.6	81.8	36.3	6.67%	47.87%	69.04%	3.4	10.1	6.9
	100	9.70%	0.00%	0.00%	70.0	228.7	107.5	19.44%	61.09%	85.95%	5.4	8.5	5.2
75%	50	73.97%	0.00%	0.00%	42.1	0.3	0.3	1.11%	39.99%	49.83%	5.0	6.7	5.0
	60	81.22%	0.00%	0.00%	47.1	1.4	1.0	6.43%	45.43%	56.86%	4.1	7.5	5.4
	70	57.25%	0.00%	0.00%	60.9	3022.7	4.9	7.50%	45.23%	80.31%	3.8	8.0	3.6
	80	54.83%	0.66%	0.00%	81.2	12941.8	39.2	4.00%	39.11%	67.44%	4.2	10.1	5.8
	90	57.37%	0.27%	0.00%	68.6	4574.8	353.5	0.00%	42.76%	68.24%	3.0	9.7	5.3
	100	46.78%	0.13%	0.00%	63.7	530.7	353.6	0.00%	48.29%	81.21%	3.8	9.7	5.2
100%	50	76.16%	6.72%	0.00%	57.3	0.9	0.8	11.33%	42.79%	63.75%	4.0	9.4	5.2
	60	56.90%	0.50%	0.00%	43.6	1.6	1.0	1.25%	46.90%	67.36%	4.1	7.5	5.6
	70	79.59%	0.10%	0.00%	54.7	124.9	6.9	0.00%	40.52%	63.08%	3.4	7.6	6.6
	80	74.20%	0.58%	0.00%	54.5	93.0	66.4	0.00%	36.94%	75.26%	3.5	8.8	5.9
	90	54.48%	0.00%	0.00%	61.6	284.0	69.7	8.33%	59.27%	80.17%	4.2	7.9	5.0
	100	81.84%	1.34%	0.00%	61.4	604.3	319.8	11.33%	56.40%	71.14%	3.4	7.8	6.8
Avg.		45.38%	0.81%	0.00%	55.35	948.69	60.35	6.68%	45.40%	64.95%	4.6	8.3	5.7

Table 4: DD set. Influence of ρ on the performances of ($\mathrm{K}_{\delta^*,\rho^*}$-QP-$kn$).

pricing objective function convex: a lower value is in principle enough. We also remark that experimentally, ρ^* is much lower than ρ^M.

We therefore perform the following experiment: we compare the bounds and CPU times yielded by three versions of our algorithms, obtained by setting either $\rho = 0$, $\rho = \bar{\rho}$ or $\rho = \rho^*$, where $\bar{\rho}$ is the *minimum* value of ρ such that the pricing objective function remains convex. The latter is found by means of a simple bisection routine in a preprocessing phase. The results on the DD set are reported in Table 4. The table includes one row for each value of n and Dens, reporting average values over the corresponding instances, and is composed by four blocks reporting, respectively, the average duality gaps (the ratio of the difference between the dual bound obtained and the optimal solution, to the optimal solution), the average CPU time spent in the column generation process, the fraction of columns in the final master problem for which the cardinality constraint is satisfied and the average number of column generation iterations to reach convergence. Each block includes three columns, one for each choice of the parameter ρ, as indicated in the leading row.

The results in Table 4 leads us to the following observations:

Experimental Observation 7. *Ensuring the pricing objective to be convex improves* *the quality of the final dual bound, but it is not enough to ensure the same high quality bounds achieved with $\rho = \rho^*$.*

Indeed, setting $\rho = 0$ leaves a huge duality gap, but also setting $\rho = \bar{\rho}$ still leaves an average duality gap of 0.81%.

Experimental Observation 8. *Choosing a high value of ρ tends to make the pricing problem harder to solve.*

In fact, the CPU time required after setting $\rho = 0$ is lower than by setting $\rho = \bar{\rho}$ or $\rho = \rho^*$.

Surprisingly, the strong bound obtained by setting $\rho = \rho^*$ can be obtained even faster than by setting $\rho = \bar{\rho}$.

Experimental Observation 9. *Decreasing ρ from ρ^* to $\bar{\rho}$ yields an increase in the total number of iterations needed to reach convergence, and therefore an overall increase in CPU time.*

Experimental Observation 10. *The higher is the value of ρ, the higher is the quantity of columns generated by the pricing algorithms that respect the cardinality constraint.*

The Experimental Observation 10 is implicitly supported by Lemma 1. However, it is fundamental to notice the following:

Experimental Observation 11. *Value ρ^* is by far* not a big-M value.

In fact when $\rho = \rho^*$ we have that, in about 35% of the times, the optimal column found by the Pricing Problem does not respect the cardinality constraint. That is, ρ^* helps to drive towards the feasibility region of the cardinality constraint without imposing to the columns generated to belong to it. This is, in our opinion, the main rationale for the good computational behaviour of our method: to partially penalize the violation of the cardinality constraint with a QCR convexification of the objective function on one side, and to complete that penalization process with the feasible region strengthening provided by DWR on the other side. This choice allows to obtain strong dual bounds without falling into the numerical troubles given by big-M penalization mechanisms.

6.4 Convexity analysis

As already mentioned, the convexification of the objective function is not needed when the instance to be solved is convex. This observation suggests that the convexity of the objective function is an aspect that should be analyzed more in detail.

In the following, we use the *fraction of positive eigenvalues* (κ) of the matrix representing the objective function to measure the convexity of an instance of (kQKP). Since we deal with problems in maximization form, an instance with a value of $\kappa = 0$ has an objective function with only negative eigenvalues and hence solving the corresponding continuous relaxation is a convex optimization problem. Instead, instances with values of $\kappa > 0$ have a nonconcave objective function, yielding to nonconvex optimization problems.

All instances of the DD set have, by construction, a value of κ approximately equal to 0.5. To better understand the impact of convexity on the performance of our algorithms, we create the following second set of instances:

- Convexity-Dependent (CD) instances.
  This set takes the 40 DD instances with $n = 50$ and modifies their objective function to obtain a set of instances with controlled values of κ. More precisely, for each original instance, we generate 11 new instances with the same values of a, b and k and with $\kappa = 0.0, 0.1, 0.2, \ldots, 0.9, 1.0$. The procedure (similar to the one described in [10]) used to obtain an objective function with a given convexity κ^* is the following:
  - We define a $n \times n$ matrix B' of real values randomly generated in the range $[-1, 1]$, and we orthonormalize it. The vectors composing B' are then used as eigenvectors of a matrix B''.
  - We randomly draw $\lceil \kappa^* n \rceil$ values in the range $[-1, 0]$ and $(n - \lceil \kappa^* n \rceil)$ values in the range $[0, 1]$, to be set as eigenvalues of B''.
  - We rebuild the matrix B'' using the spectral decomposition theorem.
  - We set $L = IB''$ and $Q = B'' - L$
  - We sort the diagonal values of L according to the order induced by the values a_j. This last reordering is done in order to have the linear objective terms correlated with the item weights.

  This procedure leads to a total of 440 instances.

In Figure 2 we synthesize the results of tests on CD instances. The values of κ are reported on the horizontal axis, while the distribution of computing times (on logarithmic scale) of the corresponding 40 instances is synthesized with a boxplot on the vertical axis. The tests led to one out of memory run for $\kappa = 0.0$ and $\kappa = 0.9$, and four for $\kappa = 1.0$; these tests are included in the plot as well.

The results in Figure 2 clearly lead to the following statement:

Experimental Observation 12. *The problem complexity tends to increase as the κ value increases.*

7 Conclusion

We presented a first comprehensive methodological and experimental study on a new family of reformulations for Binary Quadratic Problems (BQPs) based on Dantzig-Wolfe Reformulation (DWR) and Quadratic Convex Reformulation (QCR) of the objective function. In particular, we framed the new family of reformulations in a hierarchy of dominance, based on the theoretical strength of their continuous relaxation. We also tested the computational behaviour of the proposed reformulations on a large set of instances of the cardinality constrained Quadratic Knapsack Problem (kQKP) from the literature.

We were able to theoretically show that one of the new reformulation (namely ($\mathrm{K}_{\delta^*,\rho^*}$-QP-$kn$)) is able to provide the strongest dual bound. This result is experimentaly validated by the fact that such reformulation presents an integrality gap of

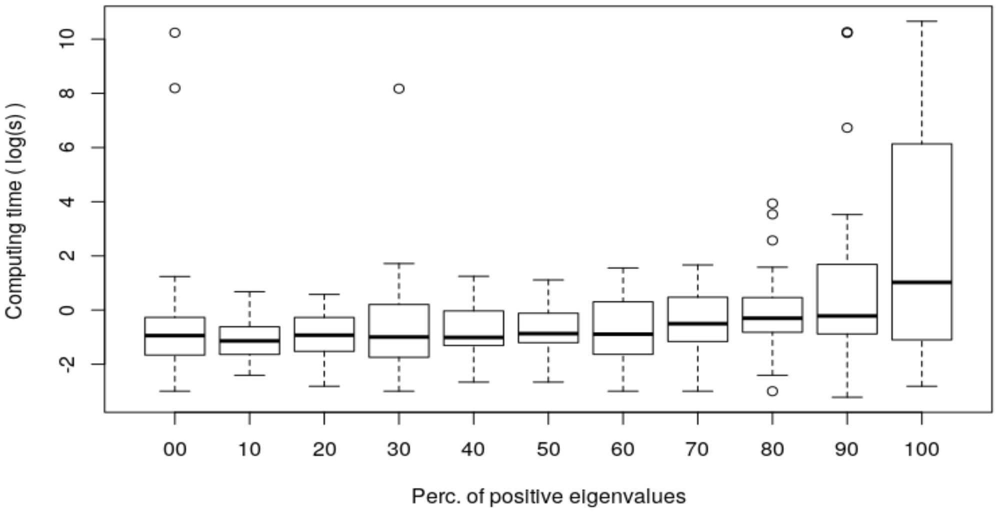

Figure 2: Distribution of computing time (log scale) of (K_{δ^*,ρ^*}-QP-kn) on CD instances.

0% for all the instances from the literature, and all the additional instances that we have created for further analyses. Moreover, such impressive bound allows to solve instances to proven integer optimality with a computing time that is lower than the computing time of a state-of-the-art commercial solver.

Our analysis of reformulations revealed that DWR and QCR are complementary in effectively handling the constraints of (K).

In particular, the set of equations used in the convexification of the objective function is of key importance because, when the convexification is applied to DWR, it implicitly increases the likelihood of generating columns respecting also such a set of equations.

In principle, our family of reformulations could be used to solve any class of BQPs. In practice, we expect them to be more effective when embeded in algorithms exploiting specific structures of the problem to solve. A promising example are mathematical formulations where equations and inequalities form blocks of well-defined substructures. Our future research will therefore focus on exploring such extensions.

References

[1] W. Adams, R. Forrester, and F. Glover. Comparisons and enhancement strategies for linearizing mixed 0-1 quadratic programs. *Discret. Optim.*, 1(2):99–120, November 2004.

[2] M. Aganagic and S. Mokhtari. Security constrained economic dispatch using nonlinear Dantzig-Wolfe decomposition. *IEEE Transactions on Power Systems*, 12(1):105–112, 1997.

[3] A. Ahlatçioglu, M. Bussieck, M. Esen, M. Guignard, J.-H. Jagla, and A. Meeraus. Combining QCR and CHR for convex quadratic pure 0-1 programming problems with linear constraints. *Annals OR*, 199(1):33–49, 2012.

[4] E. Balas and E. Zemel. An algorithm for large zero-one knapsack problems. *Operations Research*, 28:1130–1154, 1980.

[5] Saverio Basso, Alberto Ceselli, and Andrea Tettamanzi. Random sampling and machine learning to understand good decompositions. *Ann. Opera. Res.*, 284(2):501–526, 2020.

[6] M. Bergner, A. Caprara, A. Ceselli, F. Furini, M. Lübbecke, E. Malaguti, and E. Traversi. Automatic dantzig-wolfe reformulation of mixed integer programs. *Math. Program.*, 149(1-2):391–424, 2015.

[7] A. Billionnet, S. Elloumi, and A. Lambert. Extending the QCR method to general mixed-integer programs. *Mathematical Programming*, 131(1-2):381–401, 2012.

[8] A. Billionnet, S. Elloumi, and M.-C. Plateau. Improving the performance of standard solvers via a tighter convex reformulation of constrained quadratic 0-1 programs: the QCR method. *Discrete Applied Mathematics*, 157:1185–1197, 2009.

[9] A. Billionnet and E. Soutif. An exact method based on lagrangean decomposition for the 0-1 quadratic knapsak problem. *European Journal of Operational Research*, 157:565–575, 2004.

[10] C. Buchheim and A. Wiegele. Semidefinite relaxations for non-convex quadratic mixed-integer programming. *Math. Program.*, 141(1-2):435–452, 2013.

[11] A. Caprara, D. Pisinger, and P. Toth. Exact solution of the quadratic knapsack problem. *INFORMS Journal on Computing*, 11(2):125–137, 1999.

[12] A. Ceselli, L. Létocart, and E. Traversi. Dantzig wolfe decomposition and objective function convexification for binary quadratic problems: the cardinality constrained quadratic knapsack case. Technical Report 5923, 2017.

[13] GB. Dantzig and P. Wolfe. Decomposition principle for linear programs. *Operations Research*, 8:101–111, 1960.

[14] G. Desaulniers, J. Desrosiers, and M. Solomon. *Column generation*, volume 5. Springer Science & Business Media, 2006.

[15] D. Fayard and G. Plateau. Algorithm 47: an algorithm for the solution of the 0-1 knapsack problem. *Computing*, 28:269–287, 1982.

[16] A. Faye and F. Roupin. Partial lagrangian relaxation for general quadratic programming. *4OR*, 5(1):75–88, 2007.

[17] R. Fortet. Applications de l'algèbre de boole en recherche opérationnelle. *Revue Française de Recherche Opérationnelle*, 4:17–26, 1960.

[18] G. Gamrath and M. Lübbecke. Experiments with a generic Dantzig-Wolfe decomposition for integer programs. In Paola Festa, editor, *Experimental Algorithms*, pages 239–252. Springer Berlin Heidelberg, 2010.

[19] F. Glover. Improved linear integer programming formulations of nonlinear integer problems. *Management Science*, 22(4):455–460, 1975.

[20] S. Gueye and P. Michelon. “miniaturized” linearizations for quadratic 0/1 problems. *Annals of Operations Research*, 140(1):235–261, 2005.

[21] Gurobi Optimization. Gurobi, 2018. Version 8.0, http://www.gurobi.com/.

[22] K. Holmberg. Minlp: Generalized cross decomposition. In *Encyclopedia of Optimization*, pages 2148–2155. Springer, 2009.

[23] IBM. Cplex, 2018. Version 12.8.0, https://www-01.ibm.com/software/commerce/optimization/cplex-optimizer/.

[24] M. Jünger, Th.M. Liebling, D. Naddef, G. L. Nemhauser, W. R. Pulleyblank, G. Reinelt, G. Rinaldi, and L. A. Wolsey, editors. *50 Years of Integer Programming 1958-2008*. Springer, 2009.

[25] N. Krislock, J. Malick, and F. Roupin. Biqcrunch: A semidefinite branch-and-bound method for solving binary quadratic problems. *ACM Trans. Math. Softw.*, 43(4):32:1–32:23, January 2017.

[26] S. Lawphongpanich. Simplicial with truncated Dantzig-Wolfe decomposition for nonlinear multicommodity network flow problems with side constraints. *Operations Research Letters*, 26(1):33–41, 2000.

[27] C. Lemaréchal and F. Oustry. *Semidefinite relaxations and Lagrangian duality with application to combinatorial optimization*. PhD thesis, INRIA, 1999.

[28] C. Lemaréchal and A. Renaud. A geometric study of duality gaps, with applications. *Mathematical Programming*, 90(3):399–427, May 2001.

[29] L. Létocart, M.-C. Plateau, and G. Plateau. An efficient hybrid heuristic method for the 0-1 exact k-item quadratic knapsack problem. *Pesquisa Operacional*, 34(1):49–72, 2014.

[30] L. Létocart and A. Wiegele. Exact solution methods for the k-item quadratic knapsack problem. In *Lecture Notes in Computer Science*, volume 9849 of *Combinatorial Optimization*, pages 166–176. Springer, 2016.

[31] L. Liberti. Compact linearization for binary quadratic problems. *4OR*, 5(3):231–245, 2007.

[32] R. Misener and C. Floudas. GloMIQO: Global Mixed-Integer Quadratic Optimizer. *Journal of Global Optimization*, 57:3–50, 2013.

[33] D. Pisinger. The quadratic knapsack problem: a survey. *Discrete Applied Mathematics*, 155:623–648, 2007.

[34] D. Pisinger, A. B. Rasmussen, and R. Sandvik. Solution of large quadratic knapsack problems through agressive reduction. *INFORMS Journal on Computing*, 19(2):280–290, 2007.

[35] M.-C. Plateau. Quadratic convex reformulations for quadratic 0-1 programming. *4OR*, 6:187–190, 2008.

[36] F. Rendl, G. Rinaldi, and A. Wiegele. Solving Max-Cut to optimality by intersecting semidefinite and polyhedral relaxations. *Math. Programming*, 121(2):307, 2010.

[37] SCIP. Scip, 2018. Version 6.0.0, http://scip.zib.de/.

[38] H.D. Sherali and W.P. Adams. A tight linearization and an algorithm for 0-1 quadratic programming problems. *Management Science*, 32(10):1274–1290, 1986.

[39] F. Vanderbeck and M. Savelsbergh. A generic view of Dantzig-Wolfe decomposition in mixed integer programming. *Oper. Res. Lett.*, 34(3):296–306, May 2006.

[40] Philip Wolfe. A duality theorem for non-linear programming. *Quarterly of applied mathematics*, 19(3):239–244, 1961.

Dens(%)	n	(BK) Time	N. Timeout	(BK_{δ^*,ρ^*}). Time	(K_{δ^*,ρ^*}-QP-kn) Time
25	50	0.75	0	0.31	0.28
	60	0.42	0	0.19	0.36
	70	3.92	0	0.63	1.05
	80	22.92	0	1.71	7.28
	90	151.95	1	7.99	15.21
	100	280.83	0	21.49	49.89
50	50	17.03	0	0.26	0.34
	60	39.71	0	0.32	0.47
	70	33.08	4	1.73	3.16
	80	128.69	2	4.80	9.59
	90	1414.28	4	14.93	36.30
	100	1286.97	3	90.14	107.45
75	50	12.49	0	0.22	0.25
	60	14.19	1	0.59	1.04
	70	133.40	2	2.22	4.85
	80	66.53	5	9.18	39.18
	90	243.56	8	164.97	353.53
	100	145.37	8	245.92	353.63
100	50	167.46	1	0.35	0.78
	60	35.92	2	0.64	0.96
	70	11.46	6	3.52	6.92
	80	211.94	4	23.29	66.39
	90	2214.66	6	36.76	69.65
	100	116.34	5	212.84	319.84
Overall		217.58	62	35.21	60.35

Table 5: DD instances, DWR vs CPLEX

A Details on the QCR method

A.1 Semidefinite model providing δ^* and ρ^*

Let us consider ($BQ_{\delta,\rho}$) (the convexified formulation of a Binary Quadratic Problem, introduced in Section 4):

$$
\begin{aligned}
(BQ_{\delta,\rho}) \quad \min \quad & f_{\delta,\rho}(x) = x^\top Q x + L^\top x + \sum_{j\in J} \delta_j (x_j^2 - x_j) + \\
& + \rho(A_= x - b_=)^2 \\
\text{s.t.} \quad & Gx \le g \\
& Hx \le h \\
& x \in \{0,1\}^n .
\end{aligned}
$$

The SDP used to obtain the optimal parameters δ^*, ρ^* (see [8, 35]) is the following:

$$
\begin{array}{lll}
(\mathrm{SDP}_{\mathrm{BQ}_{\delta,\rho}}) & \max & \displaystyle\sum_{i=1}^{n}\sum_{j=1}^{n} Q_{ij}X_{ij} \\
& \text{s.t.} & X_{ii} = x_i \quad i = 1,\dots,n \qquad [\delta_i] \\
& & \langle A_= A_=^\top, X\rangle - 2b_=^\top A_= x = -b_=^2 \qquad [\rho] \\
& & Gx \le g \\
& & Hx \le h \\
& & \begin{pmatrix} 1 & x^t \\ x & X \end{pmatrix} \succeq 0 \\
& & x \in \mathbb{R}^n,\ X \in \mathbb{R}^{n\times n}\ .
\end{array}
$$

Namely, the optimal values δ^* and ρ^* of problem $(\mathrm{BQ}_{\delta,\rho})$ are simply given by the optimal values of the dual variables of $(\mathrm{SDP}_{\mathrm{BQ}_{\delta,\rho}})$.

A.2 Semidefinite model providing δ^*, ρ^* and Γ^*

Let us consider $(\mathrm{BQ}_{\delta,\rho,\Gamma})$:

$$
\begin{array}{lll}
(\mathrm{BQ}_{\delta,\rho,\Gamma}) & \min & f_{\delta,\rho,\Gamma}(x,z) = x^\top Qx + L^\top x + \displaystyle\sum_{j\in J}\delta_j(x_j^2 - x_j) + \rho(A_= x - b_=)^2 + \\
& & \qquad + \displaystyle\sum_{i\in J}\sum_{j\in J}\Gamma_{ij}(z_{ij} - x_i x_j) \\
& \text{s.t.} & Gx \le g \\
& & Hx \le h \\
& & z_{ij} \le x_i,\ z_{ij} \le x_j \qquad i,j = 1,\dots,n \\
& & z_{ij} \ge 0,\ z_{ij} \ge x_i + x_j - 1 \quad i,j = 1,\dots,n \\
& & x \in \{0,1\}^n\ .
\end{array}
$$

The SDP used to obtain the optimal parameters $\delta^*, \rho^*, \Gamma^*$ (see [7]) is the following:

$$
\begin{array}{lll}
(\mathrm{SDP}_{\mathrm{BQ}_{\delta,\rho,\Gamma}}) & \max & \displaystyle\sum_{i=1}^{n}\sum_{j=1}^{n} Q_{ij}X_{ij} \\
& \text{s.t.} & X_{ii} = x_i \quad i = 1,\dots,n \qquad [\delta_i] \\
& & \langle A_= A_=^\top, X\rangle - 2b_=^\top A_= x = -b_=^2 \qquad [\rho] \\
& & Gx \le g \\
& & Hx \le h \\
& & X_{ij} \le x_i,\ X_{ij} \le x_j \qquad i,j = 1,\dots,n \qquad [\Gamma_{ij}^+] \\
& & X_{ij} \ge x_i + x_j - 1,\ X_{ij} \ge 0 \quad i,j = 1,\dots,n \qquad [\Gamma_{ij}^-] \\
& & \begin{pmatrix} 1 & x^t \\ x & X \end{pmatrix} \succeq 0 \\
& & x \in \mathbb{R}^n,\ X \in \mathbb{R}^{n\times n}\ .
\end{array}
$$

Namely, the optimal values δ^*, ρ^* and Γ^* of problem $(\mathrm{BQ}_{\delta,\rho,\Gamma})$ are simply given by the optimal values of the dual variables of $(\mathrm{SDP}_{\mathrm{BQ}_{\delta,\rho,\Gamma}})$, with $\Gamma^* = \Gamma^{+*} + \Gamma^{-*}$.

B Dual of a nonlinear model reformulated with DWR

Let us consider the following formulation

$$
\begin{array}{lll}
\text{(F-NLM)} & \min & f(x) \\
& \text{s.t.} & g_i(x) \leq 0 \qquad i \in I \ [\mu_i] \\
& & x - \sum_{p \in P} x_p y^p = 0 \qquad [\pi] \\
& & \sum_{p \in P} y^p - 1 = 0 \qquad [\pi_0] \\
& & y^p \geq 0 \qquad \forall p \in \mathcal{P} \ [\nu_p]
\end{array}
$$

(we recall that the constraints $x \in [0,1]^n$ are included in the constraints $g_i(x) \leq 0$).

The formulation of its Wolfe dual is

$$
\begin{array}{lll}
\max & L(x, y, \mu, \pi, \pi_0, \nu) & \\
s.t. & \nabla_x L(x, y, \mu, \pi, \pi_0, \nu) = 0 & \\
& \nabla_y L(x, y, \mu, \pi, \pi_0, \nu) = 0 & \\
& y^p \geq 0 & \forall p \in \mathcal{P} \\
& \mu_i \geq 0 & i \in I \\
& \nu_p \leq 0 & \forall p \in \mathcal{P}
\end{array}
$$

with

$$
L(x, y, \mu, \pi, \pi_0, \nu) = f(x) + \sum_{i \in I} \mu_i g_i(x) + \pi^\top (x - \sum_{p \in \mathcal{P}} x_p y_p) + \pi_0 (\sum_{p \in \mathcal{P}} y_p - 1) + \sum_{p \in \mathcal{P}} \nu_p y_p
$$

which leads to

$$
\begin{array}{lll}
\max & L(x, y, \mu, \pi, \pi_0, \nu) & \\
\text{s.t.} & \nabla_x f(x) + \sum_{i \in I} \mu_i \nabla_x g_i(x) + \pi = 0 & \\
& -\pi^\top x_p + \pi_0 + \nu_p = 0 & \forall p \in \mathcal{P} \\
& y^p \geq 0 & \forall p \in \mathcal{P} \\
& \mu_i \geq 0 & i \in I \\
& \nu_p \leq 0 & \forall p \in \mathcal{P}
\end{array}
$$

If we substituting the definition of $\nu_p = \pi^\top x_p - \pi_0$ in $L(x, y, \mu, \pi, \pi_0, \nu)$ we have:

$$
\begin{aligned}
L(x, y, \mu, \pi, \pi_0, \nu) &= f(x) + \sum_{i \in I} \mu_i g_i(x) + \pi^\top (x - \sum_{p \in \mathcal{P}} x_p y_p) + \pi_0 (\sum_{p \in \mathcal{P}} y_p - 1) + \sum_{p \in \mathcal{P}} \nu_p y_p \\
&= f(x) + \sum_{i \in I} \mu_i g_i(x) + \pi^\top (x - \sum_{p \in \mathcal{P}} x_p y_p) + \pi_0 (\sum_{p \in \mathcal{P}} y_p - 1) + \sum_{p \in \mathcal{P}} (\pi^\top x_p - \pi_0) y_p \\
&= f(x) + \sum_{i \in I} \mu_i g_i(x) + \pi^\top (x - \sum_{p \in \mathcal{P}} x_p y_p) - \pi_0 + \sum_{p \in \mathcal{P}} (\pi^\top x_p) y_p \\
&= f(x) + \sum_{i \in I} \mu_i g_i(x) + \pi^\top x - \pi_0
\end{aligned}
$$

leading to the following dual:

$$\begin{aligned}
\max \quad & f(x) + \sum_{i\in I} \mu_i g_i(x) + \pi^\top x - \pi_0 \\
\text{s.t.} \quad & \nabla_x f(x) + \sum_{i\in I} \mu_i \nabla_x g_i(x) + \pi = 0 \\
& -\pi^\top x_p + \pi_0 \geq 0 && \forall p \in \mathcal{P} \\
& y^p \geq 0 && \forall p \in \mathcal{P} \\
& \mu_i \geq 0 && i \in I\ .
\end{aligned}$$

C Reinterpretation of the new reformulations for (BQP) with Lagrangian Duality

Let us consider the following reformulation of (BQP):

$$\begin{aligned}
(\mathrm{Q}') \quad \min \quad & w^\top S w + L^\top w \\
\text{s.t.} \quad & x_i - w_i = 0 && i = 1, \ldots, n && [\phi_i] \\
& Gw \leq g && && [\mu] \\
& Hx \leq h && && [\psi] \\
& x_i^2 - x_i = 0 && i = 1, \ldots, n && [\delta_i] \\
& (A_= x - \delta b)^2 = 0 && && [\rho] \\
& z_{ij} \leq x_i,\ \ z_{ij} \leq x_j && i, j = 1, \ldots, n && [\Xi_{ij}] \\
& z_{ij} \geq 0,\ \ z_{ij} \geq x_i + x_j - 1 && i, j = 1, \ldots, n && [\Upsilon_{ij}] \\
& z_{ij} - x_i x_j = 0 && i, j = 1, \ldots, n && [\Gamma_{ij}]
\end{aligned}$$

where for each constraint we indicate in squared brakets the corresponding Lagrangian multiplier.

Sections 3 and 4 show that the continuous relaxations of the reformulations showed in Table 1 can be viewed as a specific Lagrangian Dual of (Q′) where some of the constraints are relaxed in the Lagrangian Function, some are kept in the Lagrangian subproblem and other are dropped. For simplicitly, each constraint is represented by its Lagrangian multiplier. For example, ($\mathrm{Q}_{\delta,\rho}$-QM) is obtained by relaxing constraints $[\phi]$,$[\mu]$,$[\delta]$,$[\rho]$, keeping constraints $[\psi]$,$[\delta]$ and dropping constraints $[\Xi]$,$[\Upsilon]$,$[\Gamma]$:

$$(\mathrm{Q}_{\delta,\rho}\text{-QM}) \quad \max_{\phi,\mu,\delta,\rho} \theta(\phi,\mu,\delta,\rho) = \max_{\phi,\mu,\delta,\rho} \min_{\substack{x,w\in\mathbb{R}^n:\ Hx\leq h \\ x_i^2 - x_i = 0,\ i=1,\ldots,n}} L(x,w,\phi,\mu,\delta,\rho)$$

with $L(x,w,\phi,\mu,\delta,\rho) = w^\top Q w + L^\top w + \phi^\top(x-w) + \mu^\top(Gw-g) + \delta^\top(x^2-x) + \rho(A_= x - b_=)^2$.

D Comparing ($\mathrm{K}_{\delta^*,\rho^*}$-QP-*kn*) with CPLEX

In Table 5 we compare the time needed for computing the root node of ($\mathrm{K}_{\delta^*,\rho^*}$-QP-*kn*) (that in each test corresponds to the time needed to solve the problem to global integer optimality) with the overall time needed by `CPLEX` to solve the original formulation with general purpose techniques (BK) and (BK$_{\delta^*,\rho^*}$). For (BK) the number of instances yielding timeout is also reported: the average computing time does not include them.

The results of Table 5 provide the following observation:

Experimental Observation 13. *With (K$_{\delta^*,\rho^*}$-QP-kn) we are able to close the whole integrality gap faster than the time needed by* `CPLEX` *to solve to optimality the same instance. The time needed by (K$_{\delta^*,\rho^*}$-QP-kn) is similar to that of (BK$_{\delta^*,\rho^*}$), that is* `CPLEX` *after applying an ad-hoc convexification of the objective function.*

We remark that none of these techniques uses problem specific algorithms. In this regard, (K$_{\delta^*,\rho^*}$-QP-kn) has some additional potential, as the column generation scheme can strongly benefit from the use of ad-hoc pricing algorithms.