Copositive relaxation beats Lagrangian dual bounds in quadratically and linearly constrained quadratic optimization problems

Immanuel M. Bomze

ISOR, University of Vienna, Austria

Abstract

We study non-convex quadratic minimization problems under (possibly non-convex) quadratic and linear constraints, and characterize both Lagrangian and Semi-Lagrangian dual bounds in terms of conic optimization. While the Lagrangian dual is equivalent to the SDP relaxation (which has been known for quite a while, although the presented form, incorporating explicitly linear constraints, seems to be novel), we show that the Semi-Lagrangian dual is equivalent to a natural copositive relaxation (and this has apparently not been observed before). This way, we arrive at conic bounds tighter than the usual Lagrangian dual (and thus than the SDP) bounds. Any of the known tractable inner approximations of the copositive cone can be used for this tightening, but in particular, above mentioned characterization with explicit linear constraints is a natural, much cheaper, relaxation than the usual zero-order approximation by doubly nonnegative (DNN) matrices, and still improves upon the Lagrangian dual bounds. These approximations are based on LMIs on matrices of basically the original order plus additional linear constraints (in contrast to more familiar sum-of-squares or moment approximation hierarchies), and thus may have merits in particular for large instances where it is important to employ only a few inequality constraints (eg., n instead of $\frac{n(n-1)}{2}$ for the DNN relaxation). Further we specify sufficient conditions for tightness of the Semi-Lagrangian relaxation and show that copositivity of the slack matrix guarantees global optimality for KKT points of this problem, thus significantly improving upon a well-known second-order global optimality condition.

Key words: Copositive matrices, non-convex optimization, polynomial optimization, quadratically constrained problem, global optimality condition, approximation hierarchies

April 25, 2015

1 Introduction and basic concepts

1.1 Motivation, innovative content and organization of the paper

As is well known, the effectiveness of Lagrangian relaxation – and optimization methods in general – heavily depends on the formulation of the problem, and of the treatment of constraints. For instance, if the ground set is not the full space but rather incorporates some (simpler) constraints, we arrive at Semi-Lagrangian relaxation yielding tighter bounds than the classic Lagrangian relaxation which uses the full Euclidean space $\mathbb{R}^n$ as the ground set. However, Semi-Lagrangian dual bounds cannot always be calculated efficiently.

Here we study non-convex quadratic minimization problems under (possibly non-convex) quadratic and linear constraints, and characterize both duals in terms of conic optimization. Due to their pivotal role for applications, bounds for such type of problems receive currently much interest in the optimization community, a for sure non-exhaustive list is [2, 19, 20, 26, 28, 30, 31, 32, 34, 36, 42, 45].

In the absence of linear constraints, the full Lagrangian dual problem is equivalent to the direct semidefinite relaxation. Under additional linear constraints, we arrive at an LMI description of the Lagrangian dual which is an extension thereof, while the Semi-Lagrangian dual can be shown to result from a natural copositive relaxation. This way, we arrive at a full hierarchy of tractable conic bounds tighter than the usual Lagrangian dual (and thus than the SDP) bounds. In particular, the usual zero-order approximation by doubly nonnegative matrices improves upon the Lagrangian dual bounds. Therefore we manage a tractable approximation tightening towards Semi-Lagrangian dual bounds.

The resulting approximation hierarchy is based upon LMIs on matrices of basically the original order plus relatively few additional linear constraints, in contrast to more familiar sum-of-squares hierarchies or moment approximation hierarchies. We also relate the new relaxation with an alternative, still tighter, relaxation earlier introduced by Burer who showed that his formulation is indeed tight in an important subclass of the problem type studied here, including all mixed-binary QPs satisfying the so-called key condition. Further we study strong duality of the resulting conic problems, and also specify sufficient conditions for tightness of the Semi-Lagrangian (i.e. copositive) relaxation. We also show that copositivity of the slack matrix guarantees global optimality for KKT points of this problem. Finally, we address an alternative to replace all linear constraints by one convex quadratic. Similar aggregation approaches have been tried recently along different roads [2, 10, 23, 32].

The paper is organized as follows: first, after briefly recapitulating basic concepts, we review several variants of (Semi-)Lagrangian relaxations in the preparatory Section 2. Section 3 presents a new perspective on the full Lagrangian duals as SDPs; in Subsection 3.1, for the readers' convenience we present a summary of well-known results on all-quadratic problems without any linear constraint in a suitable context. Subsection 3.2 treats, apparently for the first time in literature, linear constraints in an explicit way and motivates the study of a cone which will serve in relaxation later on.

All these preparations will be essential in the central Section 4 where we incorporate the sign constraints into the ground set, and show that the resulting Semi-Lagrangian bounds exactly lead to the natural copositive relaxation of the all-quadratic problem with linear constraints. Next, under widely used strict feasibility conditions, we establish full strong duality of the primal-dual pair of copositive problems. However, for some formulations, strict feasibility does not hold for the original problem. Still, the major implications like primal attainability and zero duality gap for the conic relaxation can be established. Section 5 contains conditions which guarantee that the Semi-Lagrangian relaxation (and thus the copositive relaxation) is tight, and discusses global optimality conditions for a KKT point of the original problem. In Section 6 we address an alternative formulation which replaces all linear constraints by one convex quadratic, position the resulting bound to the previous natural one, and establish equivalence of this variant to Burer's relaxation which was, albeit for problems without inequality constraints, first introduced in [17]. Finally, in Section 7, we also briefly explain how to tighten Lagrangian bounds by the resulting approximation hierarchies, which may be of particular interest in large instances, i.e., in regimes where every additional linear inequality constraint "hurts" in the conic problem, forcing us to employ as few of them as possible.

The roadmap outlined above already indicates the need to somehow mix innovative contributions with novel perspectives on already known results, for presentational purposes. Therefore it may be of interest to highlight here what is new in this paper: a characterization and positioning of Semi-Lagrangian bounds within the copositive optimization framework, along with a detailed analysis of (strong) conic duality; introduction of a natural sub-zero level in approximation hierarchies, which reduces the number of linear inequality constraints to avoid memory problems in tractable relaxations; a Frank-Wolfe-type result on primal attainability of quadratic optimization problems under linear and quadratic constraints; and new second-order global optimality conditions emerging from above approach.

Summarizing, this article shall, along with a new perspective on SDP relaxations in context of more general/conic (i.e., copositive) optimization, shed new light on the question how copositivity can help in the theory and algorithmic treatment of quadratic optimization problems.

1.2 Notation and terminology

We abbreviate by $[m:n] := \{m, m+1, \dots, n\}$ the integer range between two integers m, n with $m \leq n$. By bold-faced lower-case letters we denote vectors in n-dimensional Euclidean space $\mathbb{R}^n$, by bold-faced upper case letters matrices, and by $^\top$ transposition. The positive orthant is denoted by $\mathbb{R}^n_+ := \{\mathsf{x} \in \mathbb{R}^n : x_i \geq 0 \text{ for all } i \in [1:n]\}$. I_n is the $n \times n$ identity matrix with columns e_i, $i \in [1:n]$, while $\mathsf{e} := \sum_{i=1}^{n} \mathsf{e}_i = [1, \dots, 1]^\top \in \mathbb{R}^n$ and the compact *standard simplex* is

$$\Delta := \left\{ \mathsf{x} \in \mathbb{R}^n_+ : \mathsf{e}^\top \mathsf{x} = 1 \right\} ,$$

which of course satisfies $\mathbb{R}_+ \Delta = \mathbb{R}^n_+$. The letters o and O stand for zero vectors, and zero matrices, respectively, of appropriate orders. The set of all $n \times n$ matrices is denoted by $\mathbb{R}^{n \times n}$, and the interior of a set $S \subset \mathbb{R}^n$ by S°.

For a given symmetric matrix $\mathsf{H} = \mathsf{H}^\top$, we denote the fact that H is positive-semidefinite by $\mathsf{H} \succeq \mathsf{O}$. Sometimes we write instead "H is psd." Linear forms in symmetric matrices X will play an important role in this paper; they are expressed by Frobenius duality $\langle \mathsf{S}, \mathsf{X} \rangle = \text{trace}(\mathsf{SX})$, where $\mathsf{S} = \mathsf{S}^\top$ is another symmetric matrix of the same order as X.

Given any cone $\mathcal{C}$ of symmetric $n \times n$ matrices,

$$\mathcal{C}^\star := \left\{ \mathsf{S} = \mathsf{S}^\top \in \mathbb{R}^{n \times n} : \langle \mathsf{S}, \mathsf{X} \rangle \geq 0 \text{ for all } \mathsf{X} \in \mathcal{C} \right\}$$

denotes the dual cone of $\mathcal{C}$. For instance, if $\mathcal{C} = \left\{ \mathsf{X} = \mathsf{X}^\top \in \mathbb{R}^{n \times n} : \mathsf{X} \succeq \mathsf{O} \right\}$, then $\mathcal{C}^\star = \mathcal{C}$ itself, an example of a *self-dual cone*. Trusting the sharp eyes of my readers, I chose a notation with subtle differences between the five-star denoting a dual cone, e.g., $\mathcal{C}^\star$, and the six-star, e.g. z^*, denoting optimality. Generally, (combined) subscripts will distinguish reference to various problems; e.g. $\square_{LD}$ refers to the Lagrangian dual, and $\square_S$ to a semidefinite problem. When it comes to primal-dual conic pairs, the subscripts $\square_D$ refer to the conic dual, and $\square_P$ to the primal conic problem. A subscript $\square_C$ always refers to co(mpletely)positive problems in the most frequently used form: $\square_{CD}$ indicates the dual problem over the copositive cone while $\square_{CP}$ refers to the primal problem over the completely positive cone; detailed definitions follow immediately. A subscript $\square_+$ always indicates that linear inequality constraints are treated in an explicit way. Finally, the matrix symbols Z and M are reserved for a slack matrix in the various dual conic programs, and the Shor relaxation matrix of a quadratic function, respectively.

The key notion used below is that of *copositivity*. Given a symmetric $n \times n$ matrix Q, we say that

$$\begin{aligned} \mathsf{Q} \text{ is copositive if} \quad & \mathsf{v}^\top \mathsf{Q} \mathsf{v} \geq 0 \text{ for all } \mathsf{v} \in \mathbb{R}^n_+ \,, \quad \text{and that} \\ \mathsf{Q} \text{ is strictly copositive if} \quad & \mathsf{v}^\top \mathsf{Q} \mathsf{v} > 0 \text{ for all } \mathsf{v} \in \mathbb{R}^n_+ \setminus \{\mathsf{o}\} \,. \end{aligned}$$

Strict copositivity generalizes positive-definiteness (all eigenvalues strictly positive) and copositivity generalizes positive-semidefiniteness (no eigenvalue strictly negative) of a symmetric matrix. Contrasting to positive-semidefiniteness, checking copositivity is NP-hard, see [22, 37].

The set of all copositive matrices forms a closed, convex cone, the *copositive cone*

$$\mathcal{C}^\star = \left\{ \mathsf{Q} = \mathsf{Q}^\top \in \mathbb{R}^{n \times n} : \mathsf{Q} \text{ is copositive} \right\}$$

with non-empty interior $[\mathcal{C}^\star]^\circ$ which exactly consists of all strictly copositive matrices. However, the cone $\mathcal{C}^\star$ is not self-dual. Rather one can show (denoting $s_n := n$ for $n \leq 4$ while $s_n := \frac{n(n+1)-8}{2}$ for $n \geq 5$) that $\mathcal{C}^\star$ is the dual cone of

$$\mathcal{C} = \left\{ \mathsf{X} = \mathsf{F}\mathsf{F}^\top : \mathsf{F} \text{ has } s_n \text{ columns in } \mathbb{R}^n_+ \right\} ,$$

the cone of *completely positive* matrices. Note that the factor matrix F has many more columns than rows. The upper bound s_n on the necessary number of columns was recently established by [43] and is asymptotically tight as $n \to \infty$ [14]. Anyhow, a perhaps more amenable representation is

$$\mathcal{C} = \text{conv} \, \left\{ \mathsf{x}\mathsf{x}^\top : \mathsf{x} \in \mathbb{R}^n_+ \right\} ,$$

where conv S stands for the convex hull of a set $S \subset \mathbb{R}^n$. Caratheodory's theorem then elucidates the quadratic character of s_n.

2 Lagrangian duality for quadratic problems

2.1 Different problems and different formulations

Consider two problems with quadratic constraints:

$$z^* := \inf \left\{ q_0(\mathsf{x}) : \mathsf{x} \in F \right\} \qquad \text{with } F := \left\{ \mathsf{x} \in \mathbb{R}^n : q_i(\mathsf{x}) \leq 0 \,, \, i \in [1\!:\!m] \right\} \quad (1)$$

where all $q_i(\mathsf{x}) = \mathsf{x}^\top \mathsf{Q}_i \mathsf{x} - 2\mathsf{b}_i^\top \mathsf{x} + c_i$ are quadratic functions (as the value of c_0 does not matter, we may mostly assume $c_0 = 0$, but will deviate from this in the proof of Theorem 4.2 below) for $i \in [0\!:\!m]$; and

$$z_+^* := \inf \{q_0(\mathsf{x}) : \mathsf{x} \in F \cap P\} \qquad \text{with } P := \left\{\mathsf{x} \in \mathbb{R}^n_+ : \mathsf{A}\mathsf{x} = \mathsf{a}\right\} . \qquad (2)$$

where $\mathsf{a} \in \mathbb{R}^p$ and A is a $p \times n$ matrix of full row rank p (if $P = \mathbb{R}^n_+$, i.e., $p = 0$, we will simply drop all terms involving A, a or the multipliers w introduced below). We further impose a Slater condition on the linear constraints:

$$\text{there is a point } \mathsf{y} \in P \text{ such that } y_j > 0 \text{ for all } j \in [1\!:\!n] . \qquad (3)$$

This is not customary as linear constraints do not need qualifications in the usual context; however, we will need (3) here, and it poses no restriction of generality, since we can test this condition by solving, in a preprocessing step, for all $j \in [1\!:\!n]$, the n LPs $z_j^* := \sup \{x_j : \mathsf{x} \in P\}$, and discard the variable x_j if $z_j^* = 0$. The remaining variables (we rearrange their indices again as all $j \in [1\!:\!n]$) now have the property that there is an $\mathsf{x}^{(j)} \in P$ such that $x_j^{(j)} > 0$. Taking the arithmetic mean of all $\mathsf{x}^{(j)}$ yields the desired point $\mathsf{y} \in P \cap [\mathbb{R}^+_n]^\circ$.

Neither of the optimal values z^* of (1), or z_+^* of (2) need be attained, and they could also be equal to $-\infty$ (in the unbounded case) or to $+\infty$ (in the infeasible case). Of course, we have $z^* \leq z_+^*$ due to the additional linear constraints. Considering $\mathsf{Q}_i = \mathsf{O}$ would also allow for linear inequality constraints. But it is often advisable to discriminate the functional form of constraints, and we will adhere to this principle in what follows. Therefore linear *inequality* constraints are cast into above form $\mathsf{x} \in P$ by use of slack variables, if necessary.

Note that defining $\mathsf{Q}_{m+1} = \mathsf{A}^\top \mathsf{A}$, $\mathsf{b}_{m+1} = \mathsf{A}^\top \mathsf{a}$ and $c_{m+1} = \mathsf{a}^\top \mathsf{a}$, we may rephrase the m linear constraints $\mathsf{A}\mathsf{x} = \mathsf{a}$ into one homogeneous quadratic constraint $\mathsf{z}^\top \mathsf{M}_{m+1} \mathsf{z} = \|\mathsf{A}\mathsf{x} - \mathsf{a}\|^2 = 0$. We will return later to this formulation. Still, the resulting feasible set is not of the form of F, the difference being the sign constraints $x_j \geq 0$.

Finally note that binarity constraints $x_j \in \{0, 1\}$ can be recast into two inequality constraints of the form $x_j \leq 1$ (this constraint would ensure Burer's key condition [12, 17]) and $x_j - x_j^2 \leq 0$. This fits into above formulation, but then one has to be careful with strict feasibility assumptions; also, introducing slacks for $x_j \leq 1$ will double the number of variables. We will address an alternative (Burer's relaxation) later in Subsection 6.2.

2.2 The Lagrangian (dual) functions

Now consider multipliers $\mathsf{u} \in \mathbb{R}^m_+$ of the inequality constraints $q_i(\mathsf{x}) \leq 0$, $2\mathsf{v} \in \mathbb{R}^n_+$ for the sign constraints $\mathsf{x} \in \mathbb{R}^n_+$, and $2\mathsf{w} \in \mathbb{R}^p$ for the linear equality constraints $\mathsf{A}\mathsf{x} = \mathsf{a}$ (again, the factors two are introduced for notational convenience only). Then the full Lagrangian function

$$L(\mathsf{x}; \mathsf{u}, \mathsf{v}, \mathsf{w}) := q_0(\mathsf{x}) + \sum_i u_i q_i(\mathsf{x}) - 2\mathsf{v}^\top \mathsf{x} + 2\mathsf{w}^\top(\mathsf{a} - \mathsf{A}\mathsf{x})$$

and its first two derivatives w.r.t. x are given by

$$\begin{aligned} L(\mathsf{x}; \mathsf{u}, \mathsf{v}, \mathsf{w}) &= \mathsf{x}^\top \mathsf{H}_u \mathsf{x} - 2(\mathsf{d}_\mathsf{u} + \mathsf{v} + \mathsf{A}^\top \mathsf{w})^\top \mathsf{x} + \mathsf{c}^\top \mathsf{u} + 2\mathsf{w}^\top \mathsf{a}\,, \\ \nabla_\mathsf{x} L(\mathsf{x}; \mathsf{u}, \mathsf{v}, \mathsf{w}) &= 2[\mathsf{H}_\mathsf{u} \mathsf{x} - (\mathsf{d}_\mathsf{u} + \mathsf{v} + \mathsf{A}^\top \mathsf{w})] \quad \text{and} \\ D^2_\mathsf{x} L(\mathsf{x}; \mathsf{u}, \mathsf{v}, \mathsf{w}) &= 2\mathsf{H}_\mathsf{u} \qquad \text{for all } (\mathsf{x}; \mathsf{u}, \mathsf{v}, \mathsf{w}) \in \mathbb{R}^n \times \mathbb{R}^m_+ \times \mathbb{R}^n_+ \times \mathbb{R}^p\,. \end{aligned}$$

Here we denote by $\mathsf{H}_\mathsf{u} = \mathsf{Q}_0 + \sum\limits_{i=1}^m u_i \mathsf{Q}_i$, by $\mathsf{d}_\mathsf{u} = \mathsf{b}_0 + \sum\limits_{i=1}^m u_i \mathsf{b}_i$ and by $\mathsf{c} = [c_1, \ldots, c_m]^\top$. Abbreviating $L_0(\mathsf{x}; \mathsf{u}) = L(\mathsf{x}; \mathsf{u}, \mathsf{o}, \mathsf{o})$, the Lagrangian dual function for problem (1) reads

$$\Theta_0(\mathsf{u}) := \inf \left\{ L_0(\mathsf{x}; \mathsf{u}) : \mathsf{x} \in \mathbb{R}^n \right\}\,, \tag{4}$$

and the dual optimal value is

$$z^*_{LD} := \sup \left\{ \Theta_0(\mathsf{u}) : \mathsf{u} \in \mathbb{R}^m_+ \right\}\,. \tag{5}$$

Standard weak duality implies $z^*_{LD} \leq z^*$.

The full Lagrangian dual for problem (2) with additional linear constraints reads instead

$$\Theta(\mathsf{u}, \mathsf{v}, \mathsf{w}) := \inf \left\{ L(\mathsf{x}; \mathsf{u}, \mathsf{v}, \mathsf{w}) : \mathsf{x} \in \mathbb{R}^n \right\}\,, \tag{6}$$

with dual optimal value

$$z^*_{LD,+} := \sup \left\{ \Theta(\mathsf{u}, \mathsf{v}, \mathsf{w}) : (\mathsf{u}, \mathsf{v}, \mathsf{w}) \in \mathbb{R}^m_+ \times \mathbb{R}^n_+ \times \mathbb{R}^p \right\}\,. \tag{7}$$

The idea to incorporate some of the constraints defining $F \cap P$ into the ground set, or equivalently, to relax only some of the constraints, leads to the corresponding Semi-Lagrangian (sometimes also called *partial Lagrangian*) dual and is not new, see, e.g. [26] and references therein. However, previous work has concentrated to do this with linear equality constraints, which then leads to an SDP formulation similar to those treated in the previous section. Here, we take an alternative path, incorporating the sign (i.e., inequality) constraints into the ground set, and relax all other constraints.

So we arrive at the Semi-Lagrangian variant

$$\Theta_{\text{semi}}(\mathsf{u},\mathsf{w}) := \inf\left\{L(\mathsf{x};\mathsf{u},\mathsf{o},\mathsf{w})) : \mathsf{x} \in \mathbb{R}^n_+\right\} , \tag{8}$$

with dual optimal value

$$z^*_{\text{semi}} := \sup\left\{\Theta_{\text{semi}}(\mathsf{u},\mathsf{w}) : (\mathsf{u},\mathsf{w}) \in \mathbb{R}^m_+ \times \mathbb{R}^p\right\} . \tag{9}$$

The relation between full and Semi-Lagrangian bounds is a general principle. For ease of reference, we repeat the argument here: for any $\mathsf{v} \in \mathbb{R}^n_+$,

$$\begin{aligned}
\Theta(\mathsf{u},\mathsf{v},\mathsf{w}) &= \inf\left\{L(\mathsf{x};\mathsf{u},\mathsf{v},\mathsf{w}) : \mathsf{x} \in \mathbb{R}^n\right\} \\
&\le \inf\left\{L(\mathsf{x};\mathsf{u},\mathsf{v},\mathsf{w}) : \mathsf{x} \in \mathbb{R}^n_+\right\} \\
&= \inf\left\{L(\mathsf{x};\mathsf{u},\mathsf{o},\mathsf{w}) - 2\mathsf{v}^\top\mathsf{x} : \mathsf{x} \in \mathbb{R}^n_+\right\} \\
&\le \inf\left\{L(\mathsf{x};\mathsf{u},\mathsf{o},\mathsf{w}) : \mathsf{x} \in \mathbb{R}^n_+\right\} \;=\; \Theta_{\text{semi}}(\mathsf{u},\mathsf{w}) ,
\end{aligned}$$

as $\mathsf{v}^\top\mathsf{x} \ge 0$ for all $\mathsf{x} \in \mathbb{R}^n_+$. So we arrive at the following chain of inequalities

$$z^*_{LD,+} \le z^*_{\text{semi}} \le z^*_+ ,$$

where the last inequality above follows, again, from standard weak duality.

We also have $z^*_{LD} \le z^*_{LD,+}$ as $\Theta_0(\mathsf{u}) = \Theta(\mathsf{u},\mathsf{o},\mathsf{o})$, but as z^*_{LD} and $z^*_{LD,+}$ refer to different problems, their relation cannot be seen as a tightening, but rather as a reflection of the relation $z^* \le z^*_+$ of the optimal (primal) values of (1) and (2), respectively.

2.3 Consequences of an elementary observation

We conclude this section with a key observation which is well known in the context of homogenizing polynomials, at least in the case without sign constraints. For the readers' convenience, we adapt a short proof here for the copositive case. The argument involves bordering of $n \times n$ matrices (in which context we always address the first row/column as the zeroth one. To this end, we denote by $\mathsf{e}_0 = [1, 0, \ldots, 0]^\top \in \mathbb{R}^{n+1}$, and by

$$\mathsf{J}_0 := \mathsf{e}_0\mathsf{e}_0^\top = \begin{bmatrix} 1 & \mathsf{o}^\top \\ \mathsf{o} & \mathsf{O} \end{bmatrix} .$$

Lemma 2.1 *Consider a quadratic function $q(\mathsf{x}) = \mathsf{x}^\top\mathsf{H}\mathsf{x} - 2\mathsf{d}^\top\mathsf{x} + \gamma$ defined on $\mathbb{R}^n$, with $q(\mathsf{o}) = \gamma$, $\nabla q(\mathsf{o}) = -2\mathsf{d}$ and $D^2q(\mathsf{o}) = 2\mathsf{H}$ (the factors 2 being here just for ease of later notation). Define the* Shor relaxation matrix [44]

$$\mathsf{M}(q) := \begin{bmatrix} \gamma & -\mathsf{d}^\top \\ -\mathsf{d} & \mathsf{H} \end{bmatrix} . \tag{10}$$

Then for any $\mu \in \mathbb{R}$, we have

(a) $q(\mathsf{x}) \geq \mu$ *for all* $\mathsf{x} \in \mathbb{R}^n$ *if and only if* $\mathsf{M}(q-\mu) = \mathsf{M}(q) - \mu \mathsf{J}_0 \succeq \mathsf{O}$.

(b) $q(\mathsf{x}) \geq \mu$ *for all* $\mathsf{x} \in \mathbb{R}^n_+$ *if and only if* $\mathsf{M}(q-\mu) = \mathsf{M}(q) - \mu \mathsf{J}_0 \in \mathcal{C}^\star$.

Proof. The identity $\mathsf{M}(q-\mu) = \mathsf{M}(q) - \mu \mathsf{J}_0$ is evident. Assertion (a) is proved, e.g., in [26, Lemma 1]. The argument for claim (b) is completely analogous: suppose that $q(\mathsf{x}) \geq \mu$ for all $\mathsf{x} \in \mathbb{R}^n_+$. Then H must be copositive. Indeed, otherwise consider a $\mathsf{y} \in \mathbb{R}^n_+$ such that $\mathsf{y}^\top \mathsf{H} \mathsf{y} < 0$ and look at $\mathsf{x} = t\mathsf{y}$. For large enough $t > 0$, we get

$$q(\mathsf{x}) = q(t\mathsf{y}) = t^2 \mathsf{y}^\top \mathsf{H} \mathsf{y} - 2t\mathsf{d}^\top \mathsf{y} + \gamma < \mu \,,$$

contradicting the hypothesis. So we have $[0, \mathsf{x}^\top]\mathsf{M}(q-\mu)[0, \mathsf{x}^\top]^\top = \mathsf{x}^\top \mathsf{H} \mathsf{x} \geq 0$ for all $\mathsf{x} \in \mathbb{R}^n_+$. On the other hand, we get

$$[1, \mathsf{x}^\top]\mathsf{M}(q-\mu) \begin{bmatrix} 1 \\ \mathsf{x} \end{bmatrix} = [1, \mathsf{x}^\top] \begin{bmatrix} \gamma - \mu & -\mathsf{d}^\top \\ -\mathsf{d} & \mathsf{H} \end{bmatrix} \begin{bmatrix} 1 \\ \mathsf{x} \end{bmatrix} = q(\mathsf{x}) - \mu \,, \qquad (11)$$

and the latter is nonnegative for all $\mathsf{x} \in \mathbb{R}^n_+$, by hypothesis. By homogeneity, we arrive at $\mathsf{z}^\top \mathsf{M}(q-\mu)\mathsf{z} \geq 0$ for all $\mathsf{z} \in \mathbb{R}^{n+1}_+$ and one implication is shown. The converse follows readily from (11). □

This observation implies the following identities with a duality flavor:

Corollary 2.1 *For a quadratic function* $q(\mathsf{x}) = \mathsf{x}^\top \mathsf{H} \mathsf{x} - 2\mathsf{d}^\top \mathsf{x} + \gamma$*,*

(a) $\inf \{q(\mathsf{x}) : \mathsf{x} \in \mathbb{R}^n\} = \sup \{\mu \in \mathbb{R} : \mathsf{M}(q) - \mu \mathsf{J}_0 \succeq \mathsf{O}\}$*; and*

(b) $\inf \left\{q(\mathsf{x}) : \mathsf{x} \in \mathbb{R}^n_+\right\} = \sup \{\mu \in \mathbb{R} : \mathsf{M}(q) - \mu \mathsf{J}_0 \in \mathcal{C}^\star\}$.

Note that above equalities hold, by the usual convention ($\sup \emptyset = -\infty$), also if $q(\mathsf{x})$ is unbounded from below on $\mathbb{R}^n$ or $\mathbb{R}^n_+$.

So quite naturally we are led to our first SDP, in (a), or copositive optimization problem, in (b): optimize a linear function of a variable μ under the constraint that a matrix affine-linear in μ is either psd or copositive. More generally, in a *copositive optimization problem,* for surveys see, e.g. [8, 11, 18, 24], we are given $\mathsf{r} \in \mathbb{R}^m$ as well as $m+2$ symmetric matrices $\{\mathsf{M}_0, \ldots, \mathsf{M}_m, \mathsf{J}_0\}$ of same order, and we have to maximize a linear function of $m+1$ variables $u_i \geq 0$ and $y_0 \in \mathbb{R}$ such that the affine combination $\mathsf{M}_0 - y_0 \mathsf{J}_0 + \sum_{i=1}^m u_i \mathsf{M}_i \in \mathcal{C}^\star$:

$$z^*_{CD} := \sup_{(y_0, \mathsf{u}) \in \mathbb{R} \times \mathbb{R}^m_+} \left\{ y_0 - \mathsf{r}^\top \mathsf{u} : \mathsf{M}_0 - y_0 \mathsf{J}_0 + \sum_{i=1}^m u_i \mathsf{M}_i \in \mathcal{C}^\star \right\} . \qquad (12)$$

This convex program has no local, non-global solutions, and the formulation shifts complexity from global optimization towards sheer feasibility questions (is $\mathsf{S} \in \mathcal{C}^\star$?). On the other hand, there are several hard non-convex programs which can be formulated as copositive problems, among them mixed-binary QPs or Standard QPs. The copositive formulation offers a unified view on some key classes of (mixed) continuous and discrete optimization problems. Applications range from machine learning to several combinatorial problems, including the maximum-clique problem or the maximum-cut problem.

Unlike the more popular SDP case, problem (12) is the conic dual of a problem involving a different matrix cone $\mathcal{C}$. Here we have to minimize a linear function $\langle \mathsf{M}_0, \mathsf{X}\rangle$ in a completely positive matrix variable X subject to linear constraints $\langle \mathsf{M}_i, \mathsf{X}\rangle \le r_i$, $i\in[1:m]$:

$$z^*_{CP} := \inf_{\mathsf{X}\in\mathcal{C}} \{\langle \mathsf{M}_0, \mathsf{X}\rangle : \langle \mathsf{J}_0, \mathsf{X}\rangle = 1\,,\ \langle \mathsf{M}_i, \mathsf{X}\rangle \le r_i\,,\ i\in[1:m]\}\ . \tag{13}$$

The reasons why we treat the one constraint with J_0 separately, and why we consider (13) as the primal problem, will be clear immediately.

Consider, for ease of exposition only, the all-quadratic optimization problem over the positive orthant,

$$z^*_+ := \inf\left\{q_0(\mathsf{x}) : q_i(\mathsf{x}) \le 0\,,\ i\in[1:m]\,,\ \mathsf{x}\in\mathbb{R}^n_+\right\}\ , \tag{14}$$

where all q_i are quadratic functions (resulting as a special case of (2) with empty A). Then $\mathsf{z} = [1, \mathsf{x}^\top]^\top \in \mathbb{R}^{n+1}_+$ and $\mathsf{X} = \mathsf{z}\mathsf{z}^\top$ is completely positive. Further, for $\mathsf{M}_i = \mathsf{M}(q_i)$ as defined in (10), we get $q_i(\mathsf{x}) = \mathsf{z}^\top\mathsf{M}_i\mathsf{z}$ for all $i\in[0:m]$ by (11), so we can put $\mathsf{r} = \mathsf{o}$ in (13) and (12); moreover $\langle \mathsf{J}_0, \mathsf{X}\rangle = 1$ holds. Therefore, and by weak conic duality, we get

$$z^*_{CD} \le z^*_{CP} \le z^*_+\,.$$

Strong duality for the pair (12) and (13) follows by a reasoning standard for convex problems: strict feasibility of (13) implies attainability of z^*_{CD}, and strict feasibility of (12) implies attainability of z^*_{CP}. In either of these cases we have zero duality gap, $z^*_{CD} = z^*_{CP}$. We will investigate, and formally define, strict feasibility of these conic problems in more detail in Subsection 4.3 below.

3 A new perspective on SDP relaxations

3.1 SDP and Lagrangian dual in absence of linear constraints

Dropping the sign constraints in (14), we arrive at problem (1), where, again A is empty, with its familiar SDP relaxation (see [42] in the convex case and

[28, 38, 41, 44] for nonconvex/binary variants)

$$z^*_{SD} \le z^*_{SP} \le z^* \,,$$

where

$$z^*_{SD} := \sup_{(y_0,\mathsf{u})\in\mathbb{R}\times\mathbb{R}^m_+} \left\{ y_0 : \mathsf{M}_0 - y_0\mathsf{J}_0 + \sum_{i=1}^m u_i\mathsf{M}_i \succeq \mathsf{O} \right\} \tag{15}$$

which is very similar to (12), and which is the dual of the SDP

$$z^*_{SP} := \inf_{\mathsf{X}\succeq\mathsf{O}} \left\{ \langle \mathsf{M}_0, \mathsf{X}\rangle : \langle \mathsf{J}_0, \mathsf{X}\rangle = 1,\ \langle \mathsf{M}_i, \mathsf{X}\rangle \le 0,\ i\in[1\!:\!m] \right\} \,, \tag{16}$$

the counterpart of (13). In [44] it is also shown (for the first time to the author's belief), that z^*_{SD} coincides with the Lagrangian dual for z^*.

For the readers' convenience, we start this section with a recapitulation of well-known results on all-quadratic problems without any linear constraint, put into the current context.

We have $\Theta_0(\mathsf{u}) > -\infty$ if and only if (a) $\mathsf{H}_\mathsf{u} \succeq \mathsf{O}$; and (b) the linear equation system $\mathsf{H}_\mathsf{u}\mathsf{x} = \mathsf{d}_\mathsf{u}$ has a solution. In this case $\Theta_0(\mathsf{u}) = L_0(\mathsf{x};\mathsf{u})$ for any x with $\mathsf{H}_\mathsf{u}\mathsf{x} = \mathsf{d}_\mathsf{u}$, or

$$\Theta_0(\mathsf{u}) = L_0(\mathsf{x};\mathsf{u}) = \mathsf{x}^\top\mathsf{d}_u - 2\mathsf{d}_\mathsf{u}^\top\mathsf{x} + \mathsf{c}^\top\mathsf{u} = \mathsf{c}^\top\mathsf{u} - \mathsf{d}_\mathsf{u}^\top\mathsf{x}\,.$$

So the Lagrangian dual problem can be written as a Wolfe dual with an additional psd constraint, namely as

$$z^*_{LD} = \sup\left\{ L_0(\mathsf{x};\mathsf{u}) : (\mathsf{x},\mathsf{u}) \in \mathbb{R}^n \times \mathbb{R}^m_+, \mathsf{H}_\mathsf{u} \succeq \mathsf{O}\,,\ \mathsf{H}_\mathsf{u}\mathsf{x} = \mathsf{d}_\mathsf{u} \right\} \,.$$

Unfortunately, the condition $\Theta_{\text{semi}}(\mathsf{u},\mathsf{w}) > -\infty$ does not allow for nice conditions similar to requiring $\mathsf{H}_\mathsf{u} \succeq \mathsf{O}$ and solvability of $\mathsf{H}_\mathsf{u}\mathsf{x} = \mathsf{d}_u + \mathsf{A}^\top\mathsf{w}$, which would now be the first-order condition $\nabla_\mathsf{x} L(\mathsf{x};\mathsf{u},\mathsf{o},\mathsf{w}) = \mathsf{o}$. However, for $\Theta_0(\mathsf{u})$ these conditions played a key role for the equivalence result $z^*_{LD} = z^*_{SD}$, cf. [38]. Here we will pass, also in light of the difficulties with $\Theta_{\text{semi}}(\mathsf{u},\mathsf{w})$, to a different formulation of this semidefinite relaxation for the problem (1) which immediately follows from Corollary 2.1:

Theorem 3.1 *Consider problem (1) and its Lagrangian dual function as defined in (4). Then*

$$\Theta_0(\mathsf{u}) = \sup\left\{ \mu : \mu \in \mathbb{R}\,,\ \mathsf{M}(L_0(\cdot;\mathsf{u})) - \mu\mathsf{J}_0 \succeq \mathsf{O} \right\}$$

and

$$z^*_{LD} = \sup\left\{ \mu : (\mu,\mathsf{u}) \in \mathbb{R}\times\mathbb{R}^m_+,\ \mathsf{M}(L_0(\cdot;\mathsf{u})) - \mu\mathsf{J}_0 \succeq \mathsf{O} \right\} \,.$$

Further, we have $z^*_{LD} = z^*_{SD}$ as defined in (15); so a zero duality gap $z^*_{LD} = z^*$ occurs if and only if (a) the SDP relaxation has itself no positive duality gap, and (b) the SDP relaxation is tight.

Proof. The first equation follows directly from Corollary 2.1(a), and the second equation is then immediate. But obviously

$$\mathsf{M}(L_0(\cdot;\mathsf{u})) - \mu \mathsf{J}_0 = \mathsf{M}(q_0) - y_0 \mathsf{J}_0 + \sum_{i=1}^{m} u_i \mathsf{M}(q_i)$$

when $y_0 = \mu$. Now, considering the equality constraint $\langle \mathsf{J}_0, \mathsf{X}\rangle = 1$ with multiplier $y_0 \in \mathbb{R}$ and the inequality constraints $\langle \mathsf{M}(q_i), \mathsf{X}\rangle \leq 0$ with multiplier $u_i \geq 0$, all $i \in [1:m]$, we arrive at the dual SDP (15), exactly as required. So we arrive at

$$z^*_{LD} = z^*_{SD} \leq z^*_{SP} \leq z^*$$

wherefrom the last assertion follows. □

Thus the slack matrix of the conic relaxation for (1) is

$$\mathsf{Z}(\mathsf{y}) := \mathsf{M}_0 - y_0 \mathsf{J}_0 + \sum_{i=1}^{m} u_i \mathsf{M}_i = \begin{bmatrix} \mathsf{c}^\top \mathsf{u} - y_0 & -\mathsf{d}_\mathsf{u}^\top \\ -\mathsf{d}_\mathsf{u} & \mathsf{H}_\mathsf{u} \end{bmatrix}, \tag{17}$$

where $\mathsf{y} = (y_0, \mathsf{u}) \in \mathbb{R} \times \mathbb{R}^m_+$ collects all dual variables. We will encounter updates of these slack matrices in the sequel.

3.2 Full Lagrangian dual with linear constraints

There are several, a priori different, SDP formulations for the full Lagrangian dual of (2), some adapted to special subclasses; see, e.g. [26] and references therein. If any further structural properties are missing, the formulations proposed here are general and seem to be most natural as they employ a conic constraint where the following cone $\mathcal{K}_\diamond$ occurs, which will play a significant role in terms of approximation hierarchies in Section 7 as a sub-zero level approximation of $\mathcal{C}$:

$$\mathcal{K}_\diamond := \{\mathsf{X} \text{ is psd} : X_{0j} \geq 0 \text{ for all } j \in [1:n]\} . \tag{18}$$

Its dual cone is given by

$$\mathcal{K}^\star_\diamond := \left\{ \mathsf{P} + \begin{bmatrix} 0 & \mathsf{v}^\top \\ \mathsf{v} & \mathsf{O} \end{bmatrix} : \mathsf{P} \text{ is psd}, \, \mathsf{v} \in \mathbb{R}^n_+ \right\} . \tag{19}$$

Theorem 3.2 *Consider problem (2) and its Lagrangian dual function as defined in (6). Then for all* $(\mathsf{u}, \mathsf{v}, \mathsf{w}) \in \mathbb{R}^m_+ \times \mathbb{R}^n_+ \times \mathbb{R}^p$

$$\Theta(\mathsf{u}, \mathsf{v}, \mathsf{w}) = \sup \{\mu : \mu \in \mathbb{R}, \; \mathsf{M}(L(\cdot; \mathsf{u}, \mathsf{v}, \mathsf{w})) - \mu \mathsf{J}_0 \succeq \mathsf{O}\}$$

and the full Lagrangian dual problem of (2) can be written as

$$z^*_{LD,+} = \sup\left\{\mu : (\mu, \mathsf{u}, \mathsf{w}) \in \mathbb{R} \times \mathbb{R}^m_+ \times \mathbb{R}^p,\ \mathsf{M}(L(\cdot; \mathsf{u}, \mathsf{o}, \mathsf{w})) - \mu \mathsf{J}_0 \in \mathcal{K}^\star_\diamond\right\} . \tag{20}$$

Proof. The first equation is again a direct consequence of Corollary 2.1(a). For the second, observe that

$$\mathsf{M}(L(\cdot; \mathsf{u}, \mathsf{v}, \mathsf{w})) - \mu \mathsf{J}_0 = \mathsf{M}(L(\cdot; \mathsf{u}, \mathsf{o}, \mathsf{w})) - \mu \mathsf{J}_0 - \begin{bmatrix} 0 & \mathsf{v}^\top \\ \mathsf{v} & \mathsf{O} \end{bmatrix},$$

so that $\mathsf{M}(L(\cdot; \mathsf{u}, \mathsf{o}, \mathsf{w})) - \mu \mathsf{J}_0 \in \mathcal{K}^\star_\diamond$ if and only if $\mathsf{M}(L(\cdot; \mathsf{u}, \mathsf{v}, \mathsf{w})) - \mu \mathsf{J}_0 \succeq \mathsf{O}$ for some $\mathsf{v} \in \mathbb{R}^n_+$, by (19). The result follows. □

Hence we can characterize also the full Lagrangian dual for (2) as an SDP, namely the dual of the natural SDP relaxation of (2): to this end, let us express the p linear equality constraints as $\mathsf{r}_k^\top \mathsf{x} = a_k$ with $\mathsf{r}_k \in \mathbb{R}^n$ for all $k \in [1:p]$. So $\mathsf{A}^\top = [\mathsf{r}_1, \ldots, \mathsf{r}_p]^\top$ with $\mathsf{r}_k^\top$ the kth row of A. For all $k \in [1:p]$, we define the symmetric matrices of order $n+1$

$$\mathsf{A}_k := \begin{bmatrix} 2a_k & -\mathsf{r}_k^\top \\ -\mathsf{r}_k & \mathsf{O} \end{bmatrix}. \tag{21}$$

Theorem 3.3 *For problem (2), let $\mathsf{M}_i = \mathsf{M}(q_i)$, $i \in [0:m]$, and consider the full Lagrangian dual $z^*_{LD,+}$ as defined in (7) and expressed in Theorem 3.2. Then this is the conic dual of the SDP*

$$z^*_{SP,+} := \inf\left\{\langle \mathsf{M}_0, \mathsf{X}\rangle : \mathsf{X} = \begin{bmatrix} 1 & \mathsf{x}^\top \\ \mathsf{x} & \mathsf{Y} \end{bmatrix} \succeq \mathsf{O}, \mathsf{x} \in P,\ \langle \mathsf{M}_i, \mathsf{X}\rangle \le 0, i \in [1:m]\right\} , \tag{22}$$

which can be easily seen as the natural SDP relaxation of (2). Therefore we have

$$z^*_{LD,+} = z^*_{SD,+} \le z^*_{SP,+} \le z^*_+ ,$$

and the full Lagrangian relaxation is tight, $z^*_{LD,+} = z^*_+$, if and only if (a) the SDP relaxation has zero duality gap, $z^*_{SD,+} = z^*_{SP,+}$; and (b) the primal SDP relaxation (22) is tight.

Proof. Whenever the top (zeroth) row of X reads $\mathsf{z}^\top = [1, \mathsf{x}^\top]$, we have, due to (21), $2(a_k - \mathsf{r}_k^\top \mathsf{x}) = \mathsf{z}^\top \mathsf{A}_k \mathsf{z} = \langle \mathsf{A}_k, \mathsf{X}\rangle$. Hence the constraint $\langle \mathsf{A}_k, \mathsf{X}\rangle = 0$ is equivalent to $\mathsf{r}_k^\top \mathsf{x} = a_k$. So $\mathsf{x} \in P$ is equivalent to $\mathsf{x} \in \mathbb{R}^n_+$ and $\langle \mathsf{A}_k, \mathsf{X}\rangle = 0$ for all $k \in [1:p]$. Therefore problem (22) can be alternatively written as

$$z^*_{SP,+} \quad = \inf_{\mathsf{X} \succeq \mathsf{O}, \langle \mathsf{J}_0, \mathsf{X}\rangle = 1} \left\{ \langle \mathsf{M}_0, \mathsf{X}\rangle : \begin{array}{rcll} \langle \mathsf{M}_i, \mathsf{X}\rangle & \le & 0\,, & i \in [1:m], \\ -\mathsf{e}_0^\top \mathsf{X} \mathsf{e}_j & \le & 0\,, & j \in [1:n], \\ \langle \mathsf{A}_k, \mathsf{X}\rangle & = & 0\,, & k \in [1:p] \end{array} \right\} . \tag{23}$$

Now choose multipliers $v_j \geq 0$ for the sign constraints $-\mathsf{e}_0^\top \mathsf{X} \mathsf{e}_j \leq 0$ and $w_k \in \mathbb{R}$ for the equality constraints $\langle \mathsf{A}_k, \mathsf{X} \rangle$. Then, if we dualize the SDP (22) by the standard procedure, we arrive at the new slack matrix $\mathsf{Z}_+(\mathsf{y}, \mathsf{w}) - \begin{bmatrix} 0 & \mathsf{v}^\top \\ \mathsf{v} & \mathsf{O} \end{bmatrix}$ with

$$\mathsf{Z}_+(\mathsf{y}, \mathsf{w}) := \mathsf{Z}(\mathsf{y}) + \sum_{k=1}^{p} w_k \mathsf{A}_k = \begin{bmatrix} \mathsf{c}^\top \mathsf{u} - y_0 + 2\mathsf{w}^\top \mathsf{a} & -\mathsf{d}_\mathsf{u}^\top - \mathsf{w}^\top \mathsf{A} \\ -\mathsf{d}_\mathsf{u} - \mathsf{A}^\top \mathsf{w} & \mathsf{H}_\mathsf{u} \end{bmatrix}, \quad (24)$$

where $\mathsf{Z}(\mathsf{y})$ is defined as in (17). Now notice that for $y_0 = \mu$, we have

$$\mathsf{M}(L(\cdot, \mathsf{u}, \mathsf{o}, \mathsf{w})) - \mu \mathsf{J}_0 = \mathsf{Z}_+(\mathsf{y}, \mathsf{w}) \quad \text{if } \mathsf{y} = (\mu, \mathsf{u}) \,.$$

Hence the result follows by (20) and its proof. □

3.3 Strict feasibility and strong duality for the SDP

It can easily be shown that strict feasibility of (1) implies strict feasibility of (16). Moreover, if Q_i is (strictly) positive-definite for at least one $i \in [1:m]$, then also (15) is strictly feasible, so that full strong duality holds for the primal-dual SDP pair; see [1, 38]. Under these assumptions, we arrive at

$$z_{LD}^* = z_{SD}^* = z_{SP}^* \leq z^* \,.$$

Now we pass to the problem (2) with linear constraints. By analogous reasons, if at least one Q_i is positive-definite and if there is a $\widehat{\mathsf{x}} \in P$ with $q_i(\widehat{\mathsf{x}}) < 0$ for all $i \in [1:m]$, then strong duality for the SDP pair (22) and its dual (7) holds: both optimal objective values are attained and equal the dual full Lagrangian bound, $z_{LD,+}^* = z_{SD,+}^* = z_{SP,+}^*$.

4 Semi-Lagrangian dual and copositive relaxation

4.1 A two-fold characterization of Semi-Lagrangian dual

Before we proceed to the Semi-Lagrangian case, we introduce the natural copositive relaxation of (2), in analogy to (23). Consider therefore A_k as in (21) and form the problem

$$z_{CP}^* := \inf_{\mathsf{X} \in \mathcal{C}, \langle \mathsf{J}_0, \mathsf{X} \rangle = 1} \left\{ \langle \mathsf{M}_0, \mathsf{X} \rangle : \begin{array}{rcll} \langle \mathsf{M}_i, \mathsf{X} \rangle & \leq & 0, & i \in [1:m], \\ \langle \mathsf{A}_k, \mathsf{X} \rangle & = & 0, & k \in [1:p] \end{array} \right\} \quad (25)$$

and its dual

$$z^*_{CD} \;:=\; \sup\left\{y_0 : \mathsf{Z}_+(\mathsf{y},\mathsf{w}) \in \mathcal{C}^\star, (\mathsf{y},\mathsf{w}) = (y_0,\mathsf{u},\mathsf{w}) \in \mathbb{R}\times\mathbb{R}^m_+\times\mathbb{R}^p\right\} \tag{26}$$

with the slack matrix $\mathsf{Z}_+(\mathsf{y},\mathsf{w})$ as defined in (24).

Theorem 4.1 *Consider problem (2) and its Semi-Lagrangian dual function as defined in (8), the dual z^*_{semi} as defined in (9), as well as the copositive relaxation (25) and (26). Then*

$$\Theta_{\text{semi}}(\mathsf{u},\mathsf{w}) = \sup\left\{\mu : \mu\in\mathbb{R}\,,\ \mathsf{M}(L(\cdot;\mathsf{u},\mathsf{o},\mathsf{w})) - \mu\mathsf{J}_0 \in \mathcal{C}^\star\right\}$$

and the Semi-Lagrangian dual problem of (2) can be written as

$$z^*_{\text{semi}} = \sup\left\{\mu : (\mu,\mathsf{u},\mathsf{w})\in\mathbb{R}\times\mathbb{R}^m_+\times\mathbb{R}^p,\ \mathsf{M}(L(\cdot;\mathsf{u},\mathsf{o},\mathsf{w})) - \mu\mathsf{J}_0 \in \mathcal{C}^\star\right\}\,.$$

Further, we have

$$z^*_{LD,+} \le z^*_{\text{semi}} = z^*_{CD} \le z^*_{CP} \le z^*_+\,, \tag{27}$$

and the Semi-Lagrangian relaxation is tight, $z^*_{\text{semi}} = z^*_+$, if and only if (a) the copositive relaxation has no positive duality gap, $z^*_{CD} = z^*_{CP}$, and (b) the copositive primal relaxation (25) is tight.

Proof. The first equation is now a direct consequence of Corollary 2.1(b). The remainder is as an immediate generalization of Theorem 3.3. □

So we have characterized the Semi-Lagrangian dual in two ways: (a) as the dual of the natural (primal) copositive relaxation for the problem (2); and (b) as the natural extension of the (dual) SDP relaxation for the same problem. But we can say more, in particular regarding potential computational consequences, see Section 7.

4.2 Sufficient conditions for attainability of original problem

We will proceed to develop a similar theory as in Subsection 3.3 for the copositive formulation. The aim is to replace (strict) positive-definiteness of one Q_i with strict copositivity. This is not as straightforward as it may seem at a superficial first glance, as not all relations carry over directly from the (self-dual) psd cone to the pair of dual cones $(\mathcal{C},\mathcal{C}^\star)$. For instance, from complementary slackness $\langle\mathsf{X},\mathsf{S}\rangle = 0$ it follows that the matrix product $\mathsf{XS} = \mathsf{O}$ in the SDP case but not in the copositive case.

The celebrated Frank-Wolfe theorem [27] states that any (also non-convex) quadratic function which is bounded below over a polyhedron also attains its minimum there, for a nice proof see [39]. There are many extensions, to cubic functions under the same assumptions, or to *convex* polynomial optimization problems under *convex* polynomial constraints; see, e.g. [6]. Here we deal with possibly non-convex quadratic optimization problems under (possibly non-convex) quadratic constraints. [6, p.45] presents two examples of bounded non-convex quadratics under two convex quadratic constraints where the minimum is not attained; another simple example is [39] $\inf\left\{x_1^2 : x_1 x_2 \geq 1\right\}$ with convex objective and non-convex constraint. So additional conditions are necessary to ensure this in our framework. We will now prove that strict copositivity of at least one Q_i guarantees attainability of (2), even without the assumption that the objective is bounded below on the feasible set. This result complements prior investigations [35] and a recent study [3]. Let us first establish the following auxiliary result:

Lemma 4.1 *Given arbitrary* $\mathsf{d} \in \mathbb{R}^n$ *and a symmetric* $n \times n$ *matrix* H*, consider* $q(\mathsf{x}) = \mathsf{x}^\top \mathsf{H}\mathsf{x} - 2\mathsf{d}^\top \mathsf{x}$. *For any* $\mu \in \mathbb{R}$ *define via (10)*

$$\mathsf{S}_\mu := \mathsf{M}(q) + \mu \mathsf{J}_0 = \mathsf{M}(q+\mu)\,.$$

If H *is strictly copositive, then*

(a) *there is a* $\bar{\mu} \geq 0$ *such that* S_μ *are strictly copositive for all* $\mu \geq \bar{\mu}$*;*

(b) q *is bounded from below over* $\mathbb{R}^n_+$*.*

Proof. (a) Since H is strictly copositive, $\sigma := \min\left\{\mathsf{y}^\top \mathsf{H}\mathsf{y} : \mathsf{y} \in \Delta\right\} > 0$. Further define

$$\bar{\mu} := \frac{2}{\sigma} \max\left\{(\mathsf{d}^\top \mathsf{y})^2 + 1 : \mathsf{y} \in \Delta\right\} > 0\,.$$

Now pick an arbitrary $\mathsf{z} = [x_0, \mathsf{x}^\top]^\top \in \mathbb{R}^{n+1}_+ \setminus \{\mathsf{o}\}$. If $\mathsf{x} = \mathsf{o}$, then $x_0 > 0$ and $\mathsf{z}^\top \mathsf{S}_{\bar{\mu}} \mathsf{z} = \bar{\mu} x_0^2 > 0$. If $\mathsf{x} \neq \mathsf{o}$, then $\mathsf{y} := \frac{1}{\mathsf{e}^\top \mathsf{x}} \mathsf{x} \in \Delta$ and $y_0 := \frac{1}{\mathsf{e}^\top \mathsf{x}} x_0 \geq 0$. We conclude

$$\mathsf{z}^\top \mathsf{S}_{\bar{\mu}} \mathsf{z} = (\mathsf{e}^\top \mathsf{x})^2 [\bar{\mu} y_0^2 - 2y_0 \mathsf{d}^\top \mathsf{y} + \mathsf{y}^\top \mathsf{H}\mathsf{y}] \geq (\mathsf{e}^\top \mathsf{x})^2 [\bar{\mu} y_0^2 - 2y_0 \mathsf{d}^\top \mathsf{y} + \sigma]\,.$$

Now the strictly convex function $\psi(t) = \bar{\mu} t^2 - 2(\mathsf{d}^\top \mathsf{y})t + \sigma$ attains its minimum over the positive half-ray ($t \geq 0$) either at $t = 0$ with value $\psi(0) = \sigma$, or else at $\bar{t} = \frac{\mathsf{d}^\top \mathsf{y}}{\bar{\mu}}$ with value $\psi(\bar{t}) = \sigma - \frac{(\mathsf{d}^\top \mathsf{y})^2}{\bar{\mu}} \geq \frac{\sigma}{2} > 0$. Hence

$$\mathsf{z}^\top \mathsf{S}_\mu \mathsf{z} = (\mu - \bar{\mu}) z_0^2 + \mathsf{z}^\top \mathsf{S}_{\bar{\mu}} \mathsf{z} \geq 0 + (\mathsf{e}^\top \mathsf{x})^2 \frac{\sigma}{2} > 0\,,$$

and claim (a) follows. Assertion (b) then is a consequence of (a) and Corollary 2.1(b). □

One may wonder whether there is a "weak" version of Lemma 4.1(a). However, the example $\mathsf{H} = \mathsf{O}$ and $\mathsf{d} = \mathsf{e}$ shows that S_μ is never copositive, although H is. The corresponding observation and the "strict" result for positive-(semi)definiteness is folklore, but by passing from positive-definite matrices to strictly copositive matrices, we will strengthen these findings, and also derive a stronger version of (27) in the case of linear constraints.

So let us next consider primal attainability of the original problem (2).

Theorem 4.2 *Suppose that the problem (2) is feasible, i.e., that $F \cap P \neq \emptyset$, and recall that Q_i is the Hessian of the function q_i.*

(a) *If for at least one $i \in [1:m]$ the matrix Q_i is strictly copositive, then $F \cap P$ is compact and z_+^* is attained: there is an $\mathsf{x}^* \in F \cap P$ such that $q_0(\mathsf{x}^*) = z_+^*$.*

(b) *If Q_0 is strictly copositive, then z_+^* is also attained even if $F \cap P$ is unbounded.*

Proof. For any $i \in [0:m]$ let Q_i be strictly copositive, and define the compact set $R_i := \left\{\mathsf{y} \in \Delta : \mathsf{b}_i^\top \mathsf{y} \geq 0 \text{ and } \mathsf{y}^\top(\mathsf{b}_i\mathsf{b}_i^\top - c_i\mathsf{Q}_i)\mathsf{y} \geq 0\right\}$ as well as

$$\tau_i := \max\left\{\frac{\mathsf{b}_i^\top \mathsf{y} + \sqrt{\mathsf{y}^\top(\mathsf{b}_i\mathsf{b}_i^\top - c_i\mathsf{Q}_i)\mathsf{y}}}{\mathsf{y}^\top \mathsf{Q}_i \mathsf{y}} : \mathsf{y} \in R_i\right\} < +\infty\,.$$

Consider an arbitrary $\mathsf{x} = t\mathsf{y} \in \mathbb{R}^n_+$ with $t := \mathsf{e}^\top \mathsf{x} \geq 0$ and $\mathsf{y} \in \Delta$. If now $q_i(\mathsf{x}) = t^2 \mathsf{y}^\top \mathsf{Q}_i \mathsf{y} - 2t\mathsf{b}_i^\top \mathsf{y} + c_i \leq 0$, we deduce that $\mathsf{y} \in R_i$ and that

$$t \leq \frac{\mathsf{b}_i^\top \mathsf{y} + \sqrt{\mathsf{y}^\top(\mathsf{b}_i\mathsf{b}_i^\top - c_i\mathsf{Q}_i)\mathsf{y}}}{\mathsf{y}^\top \mathsf{Q}_i \mathsf{y}}\,,$$

and hence

$$\mathsf{x} = t\mathsf{y}\,,\; t \geq 0\,,\; \mathsf{y} \in \Delta \text{ and } q_i(\mathsf{x}) \leq 0 \quad \text{imply} \quad t \leq \tau_i\,. \tag{28}$$

For $i \in [1:m]$, we deduce

$$F \cap P \subseteq \left\{\mathsf{x} \in \mathbb{R}^n_+ : q_i(\mathsf{x}) \leq 0\right\} \subseteq \left\{\mathsf{x} \in \mathbb{R}^n_+ : \mathsf{e}^\top \mathsf{x} \leq \tau_i\right\}$$

and thus z_+^* must be attained as a minimum of the continuous function q_0 over the compact set $F \cap P$. If $i = 0$, strict copositivity of the objective Hessian matrix $2\mathsf{Q}_0$ implies $z_+^* > -\infty$ by Lemma 4.1(b). Since $F \cap P \neq \emptyset$, we therefore have a finite optimal value $z_+^* \in \mathbb{R}$. Now we redefine $c_0 := -z_+^*$ and infer from (28) that $q_0(\mathsf{x}) > z_+^*$ whenever $\mathsf{e}^\top \mathsf{x} > \tau_0$ and $\mathsf{x} \in \mathbb{R}_+^n$. Therefore

$$z_+^* = \inf \{q_0(\mathsf{x}) : \mathsf{x} \in F \cap P\} = \min \left\{ q_0(\mathsf{x}) : \mathsf{x} \in F \cap P \,,\, \mathsf{e}^\top \mathsf{x} \leq \tau_0 \right\} ,$$

and the latter minimum is attained as $\left\{\mathsf{x} \in \mathbb{R}_+^n : \mathsf{e}^\top \mathsf{x} \leq \tau_0\right\}$ is compact. □

Note that an obvious modification of [35, Example 2] with $m = 2$ demonstrates the need of additional conditions: even though both Q_1 and Q_2 are psd so that the feasible region is convex (but unbounded), failure of strict copositivity of Q_0 allows for non-attainability.

4.3 Strong duality in the copositive relaxation

Now we turn to strong duality of the copositive problem.

Theorem 4.3 *Consider the copositive relaxation (25) and (26) of (2).*

(a) *Suppose that Q_i is strictly copositive for at least one $i \in [0:m]$. Then there is a $\mathsf{y} = (y_0, \mathsf{u}) \in \mathbb{R} \times \mathbb{R}_+^m$ such that $u_j > 0$ for all $j \in [1:m]$ and such that the matrix $\mathsf{Z}(\mathsf{y}) = \mathsf{Z}_+(\mathsf{y}, \mathsf{o})$ is strictly copositive, and therefore we have primal attainability and zero duality gap for the conic pair (25),(26).*

(b) *Suppose that there is an $\widehat{\mathsf{x}} \in \mathbb{R}_+^n$ such that $\mathsf{A}\widehat{\mathsf{x}} = \mathsf{a}$ and $q_i(\widehat{\mathsf{x}}) < 0$ for all $i \in [1:m]$. Then there is a matrix X in the interior of $\mathcal{C}$ such that $\langle \mathsf{J}_0, \mathsf{X} \rangle = 1$ and $\langle \mathsf{M}_i, \mathsf{X} \rangle < 0$ for all $i \in [1:m]$.*

(c) *Under the assumptions of (a) and (b), full strong duality for the primal-dual conic pair (25),(26) holds: both optimal values are attained at certain $\mathsf{X}^* \in \mathcal{C}$ and $(\mathsf{y}^*, \mathsf{w}^*) \in \mathbb{R} \times \mathbb{R}_+^m \times \mathbb{R}^p$, and there is no duality gap:*

$$z_{CD}^* = y_0^* = \langle \mathsf{M}_0, \mathsf{X}^* \rangle = z_{CP}^* \qquad \text{and} \qquad \langle \mathsf{X}^*, \mathsf{Z}_+(\mathsf{y}^*, \mathsf{w}^*) \rangle = 0 \,.$$

Proof. (a) By assumption on Q_i, the bound $\sigma := \min \left\{\mathsf{x}^\top \mathsf{Q}_i \mathsf{x} : \mathsf{x} \in \Delta\right\} > 0$. Further define

$$\alpha := \min \left\{ \sum_{j \neq i} \mathsf{x}^\top \mathsf{Q}_j \mathsf{x} : \mathsf{x} \in \Delta \right\} \in \mathbb{R}$$

and put $u_i = \max\left\{1, -\frac{2\alpha}{\sigma}\right\} > 0$. Then for all $\mathsf{x} \in \Delta$ we get by construction

$$\mathsf{x}^\top (u_i \mathsf{Q}_i + \sum_{j \neq i} \mathsf{Q}_j)\mathsf{x} \geq u_i \sigma + \alpha = \max\{-\alpha, \sigma + \alpha\} > 0 \,.$$

By positive homogeneity, we arrive at strict copositivity of the matrix $\mathsf{H}_\mathsf{u} = \mathsf{Q}_0 + \sum_{j=1}^{m} u_j \mathsf{Q}_j$ by setting $u_j := 1 > 0$ for all $j \neq i$ if $i \geq 1$, and else $u_j := \frac{1}{u_0} > 0$ if $i = 0$. By Lemma 4.1(a) and $D^2 L_0(\mathsf{x}; \mathsf{u}) = 2\mathsf{H}_\mathsf{u}$, we infer that the slack matrix $\mathsf{Z}_+(\mathsf{y}, \mathsf{o}) = \mathsf{Z}(\mathsf{y}) = \mathsf{M}(L_0(\cdot; \mathsf{u})) + \bar{t}\mathsf{J}_0$ as defined in (17) is strictly copositive for $y_0 = \mathsf{c}^\top \mathsf{u} - \bar{t}$ if $\bar{t} > 0$ is large enough.
(b) Given $\widehat{\mathsf{x}}$ as in the assumption, select $\mathsf{y} \in P \cap [\mathbb{R}^n_+]^\circ$ as in (3) and define $\mathsf{x} := (1-\varepsilon)\widehat{x} + \varepsilon \mathsf{y}$ where $\varepsilon > 0$ is chosen so small that still $q_i(\mathsf{x}) < 0$ holds for all i. This is possible by continuity of all q_i. Then $x_j > 0$ for all $j \in [1\!:\!n]$ by construction and also $\mathsf{x} \in F \cap P$. Next put $\mathsf{z} = [1, \mathsf{x}^\top]^\top$ and $\mathsf{X} = (1-\varepsilon)\mathsf{z}\mathsf{z}^\top + \varepsilon \mathsf{I}_{n+1}$. If necessary, decrease $\varepsilon > 0$ further such that still $\langle \mathsf{M}_i, \mathsf{X} \rangle < 0$ holds; again, this is possible by continuity and because

$$\langle \mathsf{M}_i, \mathsf{z}\mathsf{z}^\top \rangle = \mathsf{z}^\top \mathsf{M}_i \mathsf{z} = q_i(\mathsf{x}) < 0 \quad \text{for all } i \,.$$

Hence we can write $\mathsf{X} = [\mathsf{f}|\mathsf{B}][\mathsf{f}|\mathsf{B}]^\top$ where $\mathsf{f} = \sqrt{1-\varepsilon}\,\mathsf{z}$ has all coordinates strictly positive and $\mathsf{B} = \sqrt{\varepsilon}\,\mathsf{I}_{n+1}$ has full rank, and therefore X lies in the interior of $\mathcal{C}$ due to the improved characterization in [21]. Of course, $\langle \mathsf{J}_0, \mathsf{X} \rangle = 1$ by construction.
The remaining assertions, in particular (c), follow from Slater's theorem for convex optimization. □

Violation of the assumption in Theorem 4.3(b) will play a role in Subsection 6.1 below.

5 Tightness and second-order optimality conditions

When is the Semi-Lagrangian/copositive bound tight ?

A first answer is given by Theorem 4.1. But how is this reflected in terms of the original problem (2), i.e., of the (bordered) Hessian of the Lagrangian? Below, we will give an answer which also reveals a second-order condition sufficient for global optimality, which is weaker than the conditions derived from tightness of the Lagrangian relaxation. Note that neither F nor $F \cap P$ are, in general, convex, so strict feasibility would not imply the KKT conditions at a (local) solution, as Slater's theorem does not apply. However, tightness of the relaxations basically enforces the KKT conditions without any further constraint qualifications on (1) or on (2); in the latter case with the Semi-Lagrangian dual in a moderately generalized form though.

5.1 Recap: the full Lagrangian case, difficulty gap for SDP

Let us briefly go back to the problem (1) without linear constraints. Consider again the conditions guaranteeing strong duality for its SDP relaxation, namely (a) at least one of the Q_i is (strictly) positive-definite; and (b) there is an $\bar{\mathsf{x}} \in \mathbb{R}^n$ such that $q_i(\bar{\mathsf{x}}) < 0$ for all i. Under these conditions, [1] proved that the following two properties (a) and (b) are equivalent: (a) tightness of the semidefinite relaxation for problem (1), i.e. the equality $z^*_{SP} = z^*$; and (b) $\mathsf{Z}(q_0(\mathsf{x}^*), \mathsf{u}^*) \succeq \mathsf{O}$ for some $\mathsf{u}^* \in \mathbb{R}^m_+$ which satisfies the KKT conditions at a global solution x^* of (1).

We can say even more: if $(\bar{\mathsf{x}}, \bar{\mathsf{u}})$ is a KKT pair of (1) such that $\mathsf{H}_{\bar{\mathsf{u}}} \succeq \mathsf{O}$, then $\bar{\mathsf{x}}$ is a global solution to (1). In case of the trust region problem where $m = 1$ and $\mathsf{Q}_1 \succ \mathsf{O}$, or of a co-centered problem with two constraints where $m = 2$, $\mathsf{Q}_i \succ \mathsf{O}$ for $i \in [1:2]$ and all $\mathsf{b}_i = \mathsf{o}$, also the converse is true, so that we have always $z^*_{SP} = z^*$ in these cases, or, equivalently, for any global solution x^* there is a multiplier $\mathsf{u}^* \in \mathbb{R}^m_+$ satisfying the KKT conditions such that $\mathsf{H}_{\mathsf{u}^*} \succeq \mathsf{O}$. However, for the Celis-Dennis-Tapia (CDT) problem to minimize a nonconvex quadratic over the intersection of two ellipsoids (the inhomogeneous case of $m = 2$), the Hessian H_{u} can be indefinite at the global optimum [5] for all KKT multipliers u at x^* (generically but not always u is unique), and then there is a positive gap, $z^*_{SP} < z^*$, even though $\mathsf{Q}_i \succ \mathsf{O}$ for $i \in [1:2]$. So the converse does not hold in general, not even for problem (1) without linear constraints. For the co-centered case (general m), one has at least the Approximate S-Lemma (see [7, Lemma A6] or [29, Theorem 4.6] to bound this gap, but for the general case even this seems out of reach.

With minimal effort, one can translate above results to the full Lagrangian dual of (2), and arrive at a similar sufficient global optimality condition: if at a KKT pair $(\bar{\mathsf{x}}; \bar{\mathsf{u}}, \bar{\mathsf{v}}, \bar{\mathsf{w}})$, the slack matrix $\mathsf{Z}_+(q_0(\bar{\mathsf{x}}), \bar{\mathsf{u}}, \bar{\mathsf{w}}) \in \mathcal{K}^\star_\diamond$, then $\bar{\mathsf{x}}$ is a global solution to (2), a slight improvement over the result [31, Theorem 3.1]. The next subsection will present a much stronger result.

5.2 Semi-Lagrangian tightness and second-order optimality condition

Here, we go a step further and prove a counterpart of the above findings for the Semi-Lagrangian relaxation of problem (2). Again, this is not a straightforward generalization from positive-semidefiniteness to copositivity. In fact, we need very recent results on complementary slackness at the boundaries of $\mathcal{C}$ and $\mathcal{C}^\star$, and we need to relax the KKT conditions, too: let us say that the pair $(\mathsf{x};\mathsf{u},\mathsf{w}) \in (F\cap P)\times\mathbb{R}^m_+\times\mathbb{R}^p$ is a *generalized KKT pair* for (2) if and only if

$$\left.\begin{array}{rcll} x_j(\mathsf{H}_\mathsf{u}\mathsf{x}-\mathsf{d}_\mathsf{u}-\mathsf{A}^\top\mathsf{w})_j &=& 0 & \text{for all } j\in[1:n]\,, \\ u_i q_i(\mathsf{x}) &=& 0 & \text{for all } i\in[1:m] \text{ and} \\ w_k(a_k-\mathsf{r}_k{}^\top\mathsf{x}) &=& 0 & \text{for all } k\in[1:p]\,. \end{array}\right\} \qquad (29)$$

Let $\mathsf{v} := \mathsf{H}_\mathsf{u}\mathsf{x}-\mathsf{d}_\mathsf{u}-\mathsf{A}^\top\mathsf{w}$; then (29) is equivalent to stipulating equation $\nabla L(\mathsf{x};\mathsf{u},\mathsf{v},\mathsf{w}) = \mathsf{o}$ under the conditions $v_jx_j = 0$, $w_k(a_k-\mathsf{r}_k^\top\mathsf{x}) = 0$ and $u_iq_i(\mathsf{x}) = 0$ for all i,j,k, but without requiring $v_j \geq 0$ now.

Theorem 5.1 *Consider the following properties of problem (2):*

(a) *There is an optimal solution $\bar{\mathsf{x}}$ to (2), and for all optimal solutions x^* to (2), there is a $(\mathsf{u}^*,\mathsf{w}^*)\in\mathbb{R}^m_+\times\mathbb{R}^p$ such that $(\mathsf{x}^*;\mathsf{u}^*,\mathsf{w}^*)$ is a generalized KKT pair and such that*

$$\mathsf{Z}_+(\mathsf{y}^*,\mathsf{w}^*)\in\mathcal{C}^\star \quad \textit{for} \quad \mathsf{y}^* = (q_0(\mathsf{x}^*),\mathsf{u}^*)\,;$$

(b) *there is a global solution x^* to (2) and a $(\mathsf{u}^*,\mathsf{w}^*)\in\mathbb{R}^m_+\times\mathbb{R}^p$ such that $(\mathsf{x}^*;\mathsf{u}^*,\mathsf{w}^*)$ is a generalized KKT pair and such that*

$$\mathsf{Z}_+(\mathsf{y}^*,\mathsf{w}^*)\in\mathcal{C}^\star \quad \textit{for} \quad \mathsf{y}^* = (q_0(\mathsf{x}^*),\mathsf{u}^*)\,;$$

(c) *there is a generalized KKT pair $(\bar{\mathsf{x}};\bar{\mathsf{u}},\bar{\mathsf{w}})\in(F\cap P)\times\mathbb{R}^m_+\times\mathbb{R}^p$ such that*

$$\mathsf{Z}_+(\bar{\mathsf{y}},\bar{\mathsf{w}})\in\mathcal{C}^\star \quad \textit{for} \quad \bar{\mathsf{y}} = (q_0(\bar{\mathsf{x}}),\bar{\mathsf{u}})\,;$$

(d) *The Semi-Lagrangian relaxation is tight, $z^*_{\mathrm{semi}} = z^*_+$, and there is an optimal solution $\bar{\mathsf{x}}$ to (2).*

Then (a) $\Longrightarrow$ (b) $\Longrightarrow$ (c) $\Longrightarrow$ (d). Further, under the assumptions of Theorem 4.3(c), there is an optimal solution to (2), and all above assertions are equivalent.

Proof. The implications $(a) \Longrightarrow (b) \Longrightarrow (c)$ are obvious. To show $(c) \implies (d)$, put $\bar{y}_0 := q_0(\bar{\mathsf{x}})$, $\bar{\mathsf{y}} = [\bar{y}_0, \bar{\mathsf{u}}^\top]$, and $\bar{\mathsf{z}}^\top = [1, \bar{\mathsf{x}}^\top]$ as well as $\bar{\mathsf{X}} = \bar{\mathsf{z}}\bar{\mathsf{z}}^\top \in \mathcal{C}$. By (29), we infer $(\mathsf{d}_{\bar{\mathsf{u}}} + \mathsf{A}^\top \bar{\mathsf{w}})^\top \bar{\mathsf{x}} = \bar{\mathsf{x}}^\top \mathsf{H}_{\bar{\mathsf{u}}} \bar{\mathsf{x}}$, so that

$$\begin{aligned} \bar{y}_0 &= q_0(\bar{\mathsf{x}}) + \textstyle\sum_{i=1}^m \bar{u}_i q_i(\bar{\mathsf{x}}) + 2\sum_{k=1}^p \bar{w}_k (a_k - \mathsf{r}_k^\top \bar{\mathsf{x}}) \\ &= \mathsf{c}^\top \bar{\mathsf{u}} + 2\mathsf{a}^\top \bar{\mathsf{w}} - 2(\mathsf{d}_{\bar{\mathsf{u}}} + \mathsf{A}^\top \bar{\mathsf{w}})^\top \bar{\mathsf{x}} + \bar{\mathsf{x}}^\top \mathsf{H}_{\bar{\mathsf{u}}} \bar{\mathsf{x}}\,, \end{aligned}$$

and therefore

$$0 = (\mathsf{c}^\top \bar{\mathsf{u}} - \bar{y}_0 + 2\mathsf{a}^\top \bar{\mathsf{w}}) - 2(\mathsf{d}_{\bar{\mathsf{u}}} + \mathsf{A}^\top \bar{\mathsf{w}})^\top \bar{\mathsf{x}} + \bar{\mathsf{x}}^\top \mathsf{H}_{\bar{\mathsf{u}}} \bar{\mathsf{x}} = \bar{\mathsf{z}}^\top \mathsf{Z}_+(\bar{\mathsf{y}}, \bar{\mathsf{w}}) \bar{\mathsf{z}}\,.$$

Hence $\langle \bar{\mathsf{X}}, \mathsf{Z}_+(\bar{\mathsf{y}}, \bar{\mathsf{w}}) \rangle = \bar{\mathsf{z}}^\top \mathsf{Z}_+(\bar{\mathsf{y}}, \bar{\mathsf{w}}) \bar{\mathsf{z}} = 0$, so that $(\bar{\mathsf{X}}, \mathsf{Z}_+(\bar{\mathsf{y}}, \bar{\mathsf{w}}))$ form an optimal primal-dual pair for the copositive problem (25) and (26) with zero duality gap. We conclude

$$z_+^* \le q_0(\bar{\mathsf{x}}) = \bar{y}_0 = z_{CD}^* = z_{CP}^* = z_{\text{semi}}^* \le z_+^*$$

yielding tightness of the Semi-Lagrangian relaxation and optimality of $\bar{\mathsf{x}}$.
Now, under the assumptions of Theorem 4.3(c), there exists an optimal solution x^* to (2) by Theorem 4.2. To show that (d) implies (a), form again $\mathsf{X}^* = \mathsf{z}\mathsf{z}^\top \in \mathcal{C}$ with $\mathsf{z}^\top = [1, (\mathsf{x}^*)^\top] \in \mathbb{R}_+^{n+1}$. Then $\langle \mathsf{M}_i, \mathsf{X}^* \rangle = q_i(\mathsf{x}^*) \le 0$ for all $i \in [1\!:\!m]$ and $\langle \mathsf{J}_0, \mathsf{X}^* \rangle = 1$, so that X^* is feasible for (25). The (in)equality chain

$$z_+^* = z_{\text{semi}}^* = z_{CD}^* = z_{CP}^* \le \langle \mathsf{M}_0, \mathsf{X}^* \rangle = q_0(\mathsf{x}^*) = z_+^*$$

establishes optimality of X^*. By strong duality due to Theorem 4.3(c), there is a dual-optimal $(\mathsf{y}^*, \mathsf{w}^*) = (\mathsf{y}_0^*, \mathsf{u}^*, \mathsf{w}^*) \in \mathbb{R} \times \mathbb{R}_+^m \times \mathbb{R}^p$ such that $\mathsf{Z}_+(\mathsf{y}^*, \mathsf{w}^*) \in \mathcal{C}^\star$ and $\langle \mathsf{Z}_+(\mathsf{y}^*, \mathsf{w}^*), \mathsf{X}^* \rangle = 0$. This complementary slackness implies, at first, that

$$\left.\begin{aligned} u_i^* q_i(\mathsf{x}^*) &= u_i^* \langle \mathsf{M}_i, \mathsf{X}^* \rangle = 0 \quad \text{for all } i \in [1\!:\!m] \quad \text{and} \\ w_k^* (a_k - \mathsf{r}_k^\top \mathsf{x}^*) &= w_k^* \langle \mathsf{A}_k, \mathsf{X}^* \rangle = 0 \quad \text{for all } k \in [1\!:\!p]\,. \end{aligned}\right\} \tag{30}$$

In particular, we get $(\mathsf{a} - \mathsf{A}\mathsf{x}^*)^\top \mathsf{w}^* = \sum\limits_{k=1}^p w_k^* (a_k - {\mathsf{r}_k}^\top \mathsf{x}^*) = 0$, so that

$$\mathsf{Z}_+(\mathsf{y}^*, \mathsf{w}^*)\mathsf{X}^* = \begin{bmatrix} \mathsf{c}^\top \mathsf{u}^* - y_0^* - \mathsf{d}_{\mathsf{u}^*}^\top \mathsf{x}^* + \mathsf{a}^\top \mathsf{w}^* & [\mathsf{c}^\top \mathsf{u}^* - y_0^* - \mathsf{d}_{\mathsf{u}^*}^\top \mathsf{x}^* + \mathsf{a}^\top \mathsf{w}^*](\mathsf{x}^*)^\top \\ \mathsf{H}_{\mathsf{u}^*} \mathsf{x}^* - \mathsf{d}_{\mathsf{u}^*} - \mathsf{A}^\top \mathsf{w}^* & [\mathsf{H}_{\mathsf{u}^*} \mathsf{x}^* - \mathsf{d}_{\mathsf{u}^*} - \mathsf{A}^\top \mathsf{w}^*](\mathsf{x}^*)^\top \end{bmatrix}. \tag{31}$$

But by [43, Thm.2.1(a)] we know that $\langle \mathsf{Z}_+(\mathsf{y}^*, \mathsf{w}^*), \mathsf{X}^* \rangle = 0$ also implies $\operatorname{diag}(\mathsf{Z}_+(\mathsf{y}^*, \mathsf{w}^*)\mathsf{X}^*) = \mathsf{o}$, since $\mathsf{X}^* \in \mathcal{C}$ and $\mathsf{Z}_+(\mathsf{y}^*, \mathsf{w}^*) \in \mathcal{C}^\star$, so we infer $y_0^* = \mathsf{c}^\top \mathsf{u}^* - \mathsf{d}_{\mathsf{u}^*}^\top \mathsf{x}^* + \mathsf{a}^\top \mathsf{w}^*$ and

$$x_j^* (\mathsf{H}_{\mathsf{u}^*} \mathsf{x}^* - \mathsf{d}_{\mathsf{u}^*} - \mathsf{A}^\top \mathsf{w}^*)_j = 0 \quad \text{for all } j \in [1\!:\!n]\,. \tag{32}$$

(note that [43, Thm.2.1(b)] says that the j-th row of $\mathsf{Z}_+(\mathsf{y}^*, \mathsf{w}^*)\mathsf{X}^*$ vanishes if either $j = 0$ or if $x_j^* > 0$, which, by (31), exactly amounts to the same).

Hence $(\mathsf{x}^*; \mathsf{u}^*, \mathsf{w}^*) \in (F \cap P) \times \mathbb{R}^m_+ \times \mathbb{R}^p$ form a generalized KKT pair for (2). Now (32) also implies $(\mathsf{x}^*)^\top \mathsf{H}_{\mathsf{u}^*} \mathsf{x}^* = (\mathsf{d}_{\mathsf{u}^*} + \mathsf{A}^\top \mathsf{w}^*)^\top \mathsf{x}^*$ and therefore

$$\begin{aligned} y_0^* &= \mathsf{c}^\top \mathsf{u}^* - \mathsf{d}_{\mathsf{u}^*}^\top \mathsf{x}^* + \mathsf{a}^\top \mathsf{w}^* \\ &= \mathsf{c}^\top \mathsf{u}^* + \mathsf{a}^\top \mathsf{w}^* - \mathsf{d}_{\mathsf{u}^*}^\top \mathsf{x}^* + (\mathsf{a} - \mathsf{A}\mathsf{x}^*)^\top \mathsf{w}^* \\ &= \mathsf{c}^\top \mathsf{u}^* + 2\mathsf{a}^\top \mathsf{w}^* - (\mathsf{d}_{\mathsf{u}^*} + \mathsf{A}^\top \mathsf{w}^*)^\top \mathsf{x}^* \\ &= \mathsf{c}^\top \mathsf{u}^* + 2\mathsf{a}^\top \mathsf{w}^* - 2(\mathsf{d}_{\mathsf{u}^*} + \mathsf{A}^\top \mathsf{w}^*)^\top \mathsf{x}^* + (\mathsf{x}^*)^\top \mathsf{H}_{\mathsf{u}^*} \mathsf{x}^* \\ &= L(\mathsf{x}^*; \mathsf{u}^*, \mathsf{o}, \mathsf{w}^*) = q_0(\mathsf{x}^*) \end{aligned}$$

by (30), and assertion (a) is established. □

In fact, we have obtained the following sufficient second-order global optimality condition which need no further assumptions than stated.

Corollary 5.1 *Let $(\bar{\mathsf{x}}; \bar{\mathsf{u}}, \bar{\mathsf{w}}) \in (F \cap P) \times \mathbb{R}^m_+ \times \mathbb{R}^p$ be a generalized KKT pair for (2). If the matrix*

$$\begin{bmatrix} \mathsf{c}^\top \bar{\mathsf{u}} + 2\mathsf{a}^\top \bar{\mathsf{w}} - q_0(\bar{\mathsf{x}}), & -(\mathsf{d}_{\bar{\mathsf{u}}} + \mathsf{A}^\top \bar{\mathsf{w}})^\top \\ -(\mathsf{d}_{\bar{\mathsf{u}}} + \mathsf{A}^\top \bar{\mathsf{w}}), & \mathsf{H}_{\bar{\mathsf{u}}} \end{bmatrix} \tag{33}$$

is copositive, then $\bar{\mathsf{x}}$ is a global solution to (2).

Proof. Observe that in the proof of $(c) \Rightarrow (d)$ of Theorem 5.1 above, we never used one of the conditions in Theorem 4.3. So regardless of these, global optimality of $\bar{\mathsf{x}}$ holds, along with tightness and zero duality gap, $z^*_{\text{semi}} = z^*_+ = z^*_{CP} = z^*_{CD} = q_0(\bar{\mathsf{x}})$. □

The significance of above result is that it considerably tightens previously known second-order sufficient global optimality conditions; for the role of copositivity in second-order optimality conditions for general smooth optimization problems, refer to [9]. While checking copositivity is NP-hard, the slack matrix may lie in a slightly smaller but tractable approximation cone (cf. Section 7 below), and then global optimality is guaranteed even in cases where the slack matrix is indefinite.

Problem (2) may have many (generalized) KKT points $\bar{\mathsf{x}}$, some of which can be detected with not too much effort by local optimization procedures; cf. [45]. Next, we may solve the linear equations for $(\bar{\mathsf{u}}, \bar{\mathsf{w}})$, and then test a sufficient copositivity criterion for the matrix in (33), to get a certificate for global optimality of $\bar{\mathsf{x}}$. The condition is weaker than that addressed at the end of Subsection 5.1 in two aspects: it deals with *generalized* KKT pairs, and it requires only $Z_+(\bar{\mathsf{y}}, \bar{\mathsf{w}}) \in \mathcal{C}^\star$ rather than $Z_+(\bar{\mathsf{y}}, \bar{\mathsf{w}}) \in \mathcal{K}_\diamond^\star$. Recall that the sub-zero level approximation cone $\mathcal{K}_\diamond^\star$ is much smaller than $\mathcal{C}^\star$.

The difference can also be expressed in properties of the Hessian $\mathsf{H}_{\bar{\mathsf{u}}}$ of the Lagrangian: indeed, the condition $Z_+(\bar{\mathsf{y}}, \bar{\mathsf{w}}) \in \mathcal{K}_\diamond^\star$ (giving tightness $z^*_{LD,+} = z^*_+$) implies that its lower right principal submatrix $\mathsf{H}_{\bar{\mathsf{u}}}$ has to be psd, and we know this is too strong in some cases (recall Subsection 5.1), whereas $Z_+(\bar{\mathsf{y}}, \bar{\mathsf{w}}) \in \mathcal{C}^\star$ (giving tightness $z^*_{\text{semi}} = z^*_+$), by the same argument, only yields copositivity of $\mathsf{H}_{\bar{\mathsf{u}}}$. Of course, this happens with higher frequency than positive-definiteness of the Hessian, and the discrepancy is not negligible, see [13] for a related simulation study.

Example. For any n, consider an indefinite, but copositive matrix Q_0 (e.g., $\mathsf{Q}_0 = \mathsf{e}\mathsf{e}^\top - \frac{1}{2}\mathsf{I}_n$). Further suppose that the origin o is feasible w.r.t. the quadratic constraints, i.e. $q_i(\mathsf{o}) \leq 0$ for all $i \in [1:m]$. Here q_i are (for ease of exposition assumed to be) concave quadratic constraint functions of arbitrary number m. Evidently, o is a critical point of the objective $q_0(\mathsf{x}) := \mathsf{x}^\top \mathsf{Q}_0 \mathsf{x}$ and so $(\mathsf{o}; \mathsf{o})$ is a KKT (in fact, optimal) pair of the problem

$$z^*_+ := \min \left\{ q_0(\mathsf{x}) : q_i(\mathsf{x}) \leq 0\,,\ i \in [1:m]\,,\ \mathsf{x} \in \mathbb{R}^n_+ \right\}\,.$$

However,

$$\mathsf{Z}_+(y_0, \mathsf{o}) = \begin{bmatrix} -y_0 & \mathsf{o}^\top \\ \mathsf{o} & \mathsf{Q}_0 \end{bmatrix} \notin \mathcal{K}_\diamond^\star \quad \text{for all } y_0 \in \mathbb{R}\,,$$

because Q_0 is indefinite. Moreover, for all $\mathsf{y} = (y_0, \mathsf{u}) \in \mathbb{R} \times \mathbb{R}^m_+$, we have $\mathsf{Z}_+(\mathsf{y}) \notin \mathcal{K}_\diamond^\star$ for a similar reason: for no $\mathsf{u} \in \mathbb{R}^m_+$, the block H_u can be positive-semidefinite. Therefore there is a Lagrangian relaxation gap, $z^*_{LD,+} = -\infty < 0 = z^*_+$ while the semi-Lagrangian gap is closed; indeed,

$$z^*_{CD} = \sup \left\{ y_0 : \begin{bmatrix} -y_0 & \mathsf{o}^\top \\ \mathsf{o} & \mathsf{Q}_0 \end{bmatrix} \in \mathcal{C}^\star \right\} = 0 = z^*_+\,.$$

If constraints q_i are chosen instead such that $F \subseteq \hat{\mathsf{x}} + \mathbb{R}^n_+$ for some $\hat{\mathsf{x}} \in [\mathbb{R}^n_+]^\circ$ with $q_i(\hat{\mathsf{x}}) = 0$, rendering some or all quadratic constraints binding, and some or all linear ones non-binding, we can have the same effect with $\hat{\mathsf{x}}$ instead of o by shifting the objective: $q_0(\mathsf{x}) = (\mathsf{x} - \hat{\mathsf{x}})^\top \mathsf{Q}_0 (\mathsf{x} - \hat{\mathsf{x}})$, as the second-order properties remain unaffected by these changes.

6 Alternative copositive relaxations: aggregation and Burer's approach coincide

6.1 Replacing all linear constraints by one quadratic

Next let us replace the p linear constraints $\mathsf{Ax} = \mathsf{a}$ by one quadratic constraint $q_{m+1}(\mathsf{x}) := \|\mathsf{Ax} - \mathsf{a}\|^2 = 0$, corresponding to

$$\mathsf{M}_{m+1} = \mathsf{M}(q_{m+1}) = \begin{bmatrix} \mathsf{a}^\top \mathsf{a} & -a^\top \mathsf{A} \\ -\mathsf{A}^\top \mathsf{a} & \mathsf{A}^\top \mathsf{A} \end{bmatrix} .$$

Of course, we cannot expect full strong duality for the original copositive formulation (25), and neither for the more accurate version, namely the copositive representation of the Semi-Lagrangian dual of this alternative:

$$\left.\begin{array}{l} z^*_{CP,\mathrm{agg}} := \inf\limits_{\mathsf{X}\in\mathcal{C}} \{\langle \mathsf{M}_0, \mathsf{X}\rangle : \langle \mathsf{M}_i, \mathsf{X}\rangle \le 0, \langle \mathsf{J}_0, \mathsf{X}\rangle = 1, \langle \mathsf{M}_{m+1}, \mathsf{X}\rangle = 0\} \\ z^*_{CD,\mathrm{agg}} := \sup\left\{y_0 : \mathsf{Z}_{\mathrm{agg}}(\bar{\mathsf{y}}) \in \mathcal{C}^\star, \bar{\mathsf{y}} = [y_0, \mathsf{u}^\top, u_{m+1}]^\top \in \mathbb{R}\times\mathbb{R}^m_+ \times \mathbb{R}\right\} , \end{array}\right\} , \tag{34}$$

where $\mathsf{Z}_{\mathrm{agg}}(\bar{\mathsf{y}}) := \mathsf{Z}(y_0, \mathsf{u}) + u_{m+1}\mathsf{M}_{m+1}$ and Z is defined in (17). Obviously, we have

$$z^*_{SD,\mathrm{agg}} \le z^*_{CD,\mathrm{agg}} \le z^*_{CP,\mathrm{agg}} \le z^*_+ \quad \text{and also} \quad z^*_{SD,\mathrm{agg}} \le z^*_{SP,\mathrm{agg}} \le z^*_{CP,\mathrm{agg}} ,$$

if we consider the sub-zero level relaxations

$$\left.\begin{array}{l} z^*_{SP,\mathrm{agg}} := \inf\limits_{\mathsf{X}\in\mathcal{K}_\diamond} \{\langle \mathsf{M}_0, \mathsf{X}\rangle : \langle \mathsf{M}_i, \mathsf{X}\rangle \le 0, \langle \mathsf{J}_0, \mathsf{X}\rangle = 1, \langle \mathsf{M}_{m+1}, \mathsf{X}\rangle = 0\} \\ z^*_{SD,\mathrm{agg}} := \sup\left\{y_0 : \mathsf{Z}_{\mathrm{agg}}(\bar{\mathsf{y}}) \in \mathcal{K}^\star_\diamond, \bar{\mathsf{y}} = [y_0, \mathsf{u}^\top, u_{m+1}]^\top \in \mathbb{R}\times\mathbb{R}^m_+ \times \mathbb{R}\right\} \end{array}\right\} , \tag{35}$$

where the primal $z^*_{SP,\mathrm{agg}}$ tightens the Lagrangian relaxation as is shown below (the author is indebted to a diligent referee for this hint):

Theorem 6.1 *Above primal sub-zero level relaxation tightens the gap from its counterpart in (22):*

$$z^*_{LD,+} = z^*_{SD,+} \le z^*_{SP,+} \le z^*_{SP,\mathrm{agg}} .$$

Proof. Only the rightmost inequality above needs a proof. Let

$$\mathsf{X} = \begin{bmatrix} 1 & \mathsf{x}^\top \\ \mathsf{x} & \mathsf{Y} \end{bmatrix} \quad \text{be (35)-feasible.}$$

Since $\mathsf{X} \succeq \mathsf{O}$ by $\mathsf{X} \in \mathcal{K}_\diamond$, we have $\mathsf{Y} \succeq \mathsf{xx}^\top$. As $\mathsf{A}^\top\mathsf{A} \succeq \mathsf{O}$, we have also $\|\mathsf{Ax}\|^2 = \mathsf{x}^\top\mathsf{A}^\top\mathsf{Ax} \le \langle \mathsf{A}^\top\mathsf{A}, \mathsf{Y}\rangle$, entailing

$$\begin{array}{rcl} \|\mathsf{Ax} - \mathsf{a}\|^2 & = & \|\mathsf{Ax}\|^2 - 2\mathsf{a}^\top\mathsf{Ax} + \|\mathsf{a}\|^2 \\ & \le & \langle \mathsf{A}^\top\mathsf{A}, \mathsf{Y}\rangle - 2\mathsf{a}^\top\mathsf{Ax} + \|\mathsf{a}\|^2 = \langle \mathsf{M}_{m+1}, \mathsf{X}\rangle . \end{array}$$

Now $\mathsf{X} \in \mathcal{K}_\diamond$ yields also $\mathsf{x} \in \mathbb{R}^n_+$, so that $\langle \mathsf{M}_{m+1}, \mathsf{X} \rangle = 0$ implies, by above, finally $\mathsf{x} \in P$. Hence X is also (22)-feasible, and the inequality follows. □

Evidently, for no x we can have $q_{m+1}(\mathsf{x}) < 0$. Still we have zero duality gap and primal attainability for the conic pairs, if problem (2) is feasible at all, under mild conditions:

Theorem 6.2 *Consider the case $q_{m+1}(\mathsf{x}) = \|\mathsf{A}\mathsf{x} - \mathsf{a}\|^2$. Suppose that at least one Q_i is strictly copositive for $i \in [0\!:\!m+1]$ (note that $\mathsf{Q}_{m+1} = \mathsf{A}^\top \mathsf{A}$ is so if and only if ker $\mathsf{A} \cap \mathbb{R}^n_+ = \{\mathsf{o}\}$). Then both primal/dual conic pairs, (25)/(26) and (34), have zero duality gap and the primal optimal value is attained if there is an $\bar{\mathsf{x}} \in F \cap P$:*
for some $\mathsf{X}^* \in \mathcal{C}$ such that $\langle \mathsf{M}_i, \mathsf{X}^* \rangle \leq 0$ for all $i \in [1\!:\!m]$ as well as $\langle \mathsf{J}_0, \mathsf{X}^* \rangle = 1$ and $\langle \mathsf{M}_{m+1}, \mathsf{X}^* \rangle = 0$, we have

$$z^*_{CD,\mathrm{agg}} = z^*_{CP,\mathrm{agg}} = \langle \mathsf{M}_0, \mathsf{X}^* \rangle \,.$$

Proof. First note that the primal problem in (34) is feasible since $\mathsf{X} = \mathsf{z}\mathsf{z}^\top$ with $\mathsf{z}^\top = [1, \bar{\mathsf{x}}^\top]$ satisfies all constraints. Next construct a strictly feasible $\mathsf{Z}_{\mathrm{agg}}(\bar{\mathsf{y}}) = \mathsf{Z}(\mathsf{y})$ with $u_{m+1} = 0$ from $\mathsf{Z}(\mathsf{y})$ as in the proof of Theorem 4.3(a). Now the result follows from Slater's principle, applied to the conic primal/dual pair. □

6.2 Burer's relaxation and aggregation

We now pass to an alternative put forward by Burer in his seminal paper [17], although this is not made explicit there in full generality; but see the more recent papers [19, 20]. Basically, he concentrated on mixed-binary, linearly constrained quadratic optimization problems, but extended the results to problems with additional quadratic *equality* constraints, e.g., complementarity constraints. The focus of [17] was laid on reformulation rather than on relaxation, and the problem (2) with *inequality* constraints was not treated there. However the approach in [17] can be easily extended to general quadratic inequality constraints, namely to complement the condition $\langle \mathsf{A}_k, \mathsf{X} \rangle = 0$ by another one resulting from squaring the linear constraint $\mathsf{r}_k^\top \mathsf{x} = a_k$: again, with $\mathsf{X} = [1, \mathsf{x}^\top]^\top [1, \mathsf{x}^\top]$, we have

$$\langle \mathsf{r}_k \mathsf{r}_k^\top, \mathsf{x}\mathsf{x}^\top \rangle = (\mathsf{r}_k^\top \mathsf{x})^2 = a_k^2 \quad \iff \quad \langle \mathsf{B}_k, \mathsf{X} \rangle = 0 \text{ with } \mathsf{B}_k := \begin{bmatrix} -a_k^2 & \mathsf{o}^\top \\ \mathsf{o} & \mathsf{r}_k \mathsf{r}_k^\top \end{bmatrix} .$$

So we arrive at another copositive relaxation for (2),

$$\left.\begin{array}{rcl} z^*_{CP,\mathrm{Burer}} &:=& \inf\limits_{\mathsf{X}\in\mathcal{C}} \left\{ \langle \mathsf{M}_0, \mathsf{X}\rangle : \begin{array}{rcll} \langle \mathsf{M}_i, \mathsf{X}\rangle & \le & 0\,, & i\in[1:m],\\ \langle \mathsf{A}_k, \mathsf{X}\rangle & = & 0\,, & k\in[1:p],\\ \langle \mathsf{B}_k, \mathsf{X}\rangle & = & 0\,, & k\in[1:p],\\ \langle \mathsf{J}_0, \mathsf{X}\rangle & = & 1 & \end{array} \right\} \text{ and} \\ z^*_{CD,\mathrm{Burer}} &:=& \sup \left\{ y_0 : \begin{array}{l} \mathsf{y} = (y_0, \mathsf{u}) \in \mathbb{R}\times\mathbb{R}^m_+\,,\\ (\mathsf{w},\mathsf{z}) \in \mathbb{R}^p\times\mathbb{R}^p\,,\\ \mathsf{Z}_{\mathrm{Burer}}(\mathsf{y},\mathsf{w},\mathsf{z}) \in \mathcal{C}^\star \end{array} \right\} \end{array}\right\} \qquad (36)$$

with $\mathsf{Z}_{\mathrm{Burer}}(\mathsf{y},\mathsf{w},\mathsf{z}) = \mathsf{Z}_+(\mathsf{y},\mathsf{w}) + \sum_{k=1}^{p} z_k \mathsf{B}_k$, which is what we refer to as *Burer's (copositive) relaxation* in our current context. Since $\mathsf{Z}_{\mathrm{Burer}}(\mathsf{y},\mathsf{w},\mathsf{o}) = \mathsf{Z}_+(\mathsf{y},\mathsf{w})$, we get

$$z^*_{\mathrm{semi}} = z^*_{CD} \le z^*_{CD,\mathrm{Burer}} \le z^*_{CP,\mathrm{Burer}} \le z^*_+$$

and similarly $z^*_{CP} \le z^*_{CP,\mathrm{Burer}} \le z^*_+$. As with (34) and (35), there is a sub-zero approximation variant where $(\mathcal{C}^\star, \mathcal{C})$ in (36) is replaced with $(\mathcal{K}^\star_\diamond, \mathcal{K}_\diamond)$. The optimal values will be referred to as $z^*_{SD,\mathrm{Burer}}$ and $z^*_{SP,\mathrm{Burer}}$, respectively.

For linearly constrained quadratic problems with binarity constraints which are formulated as $q_j(\mathsf{x}) = x_j - x_j^2 = 0$ (and relaxed as $\langle \mathsf{M}(q_j), \mathsf{X}\rangle = 0$ with multipliers $u_j \in \mathbb{R}$), the duality gap for this copositive relaxation is zero. Indeed, for $\mathsf{u} = t\mathsf{e}$ and $\mathsf{y} = (y_0, \mathsf{u})$,

$$\mathsf{Z}_{\mathrm{Burer}}(\mathsf{y},\mathsf{o},\mathsf{o}) = \left[\begin{array}{cc} -y_0 & (t\mathsf{e}-\mathsf{b}_0)^\top \\ (t\mathsf{e}-\mathsf{b}_0) & -2t\mathsf{I}_n + \mathsf{Q}_0 \end{array}\right]$$

can always be made strictly copositive in light of Lemma 4.1 above, e.g. for $t = \min\{3\lambda_{\min}(\mathsf{Q}_0), -1\}$. Decreasing also y_0 if necessary, we even can achieve $\mathsf{Z}_{\mathrm{Burer}}(\mathsf{y},\mathsf{o},\mathsf{o}) \in [\mathcal{K}^\star_\diamond]^\circ$.

Observe that in this case, no sign restrictions to u apply, and that, as with the aggregated formulation, strict primal feasibility cannot be inferred by the general arguments in Theorem 4.3(b). For this type of problems (and for the extension to some quadratic equality constraints), Burer showed in [17] that under a mild condition, this relaxation is always tight, $z^*_{CD,\mathrm{Burer}} = z^*_{CP,\mathrm{Burer}} = z^*_+$.

Let us return to the general case with additional quadratic inequality constraints where a positive relaxation gap $z^*_{CP,\mathrm{Burer}} < z^*_+$ cannot be excluded. We now show that aggregation and Burer's relaxation essentially coincide, both for the exact and for the approximate variant:

Theorem 6.3 *In the primal, Burer's relaxation is equivalent to the aggregation one, and it (weakly) tightens the dual one; the same relations hold at sub-zero level of approximation:*

$$\left.\begin{array}{ccccccc} z^*_{CD,\mathrm{agg}} & \le & z^*_{CD,\mathrm{Burer}} & \text{and} & z^*_{CP,\mathrm{Burer}} & = & z^*_{CP,\mathrm{agg}}\,, \\ z^*_{SD,\mathrm{agg}} & \le & z^*_{SD,\mathrm{Burer}} & \text{and} & z^*_{SP,\mathrm{Burer}} & = & z^*_{SP,\mathrm{agg}}\,. \end{array}\right\} \tag{37}$$

Further, in the case of zero conic duality gap of the aggregated version, the first four of these bounds coincide, and likewise the last four ones:

$$\left.\begin{array}{ccccccccc} z^*_{CD,\mathrm{agg}} & = & z^*_{CD,\mathrm{Burer}} & = & z^*_{CP,\mathrm{Burer}} & = & z^*_{CP,\mathrm{agg}} & \text{and} \\ z^*_{SD,\mathrm{agg}} & = & z^*_{SD,\mathrm{Burer}} & = & z^*_{SP,\mathrm{Burer}} & = & z^*_{SP,\mathrm{agg}}\,. & \end{array}\right\} \tag{38}$$

Proof. Let us start with the observation that

$$\mathsf{C}_k := a_k\mathsf{A}_k + \mathsf{B}_k = \begin{bmatrix} a_k^2 & -a_k\mathsf{r}_k^\top \\ -a_k\mathsf{r}_k & \mathsf{r}_k\mathsf{r}_k^\top \end{bmatrix} = [a_k, -\mathsf{r}_k^\top]^\top[a_k, -\mathsf{r}_k^\top] \succeq \mathsf{O}\,.$$

Hence all C_k are psd., so for any $\mathsf{X} \in \mathcal{K}_\diamond$, the conditions $\langle \mathsf{C}_k, \mathsf{X}\rangle = 0$ for all $k\in[1:p]$ are equivalent to

$$\sum_{k=1}^{p}\langle \mathsf{C}_k, \mathsf{X}\rangle = 0\,,$$

i.e., to a single homogeneous linear constraint. But

$$\sum_{k=1}^{p}\mathsf{C}_k = \begin{bmatrix} \mathsf{a}^\top\mathsf{a} & -\mathsf{a}^\top\mathsf{A} \\ -\mathsf{A}^\top\mathsf{a} & \mathsf{A}^\top\mathsf{A} \end{bmatrix} = \mathsf{M}_{m+1}\,,$$

so that the constraint $\langle \mathsf{M}_{m+1}, \mathsf{X}\rangle = 0$ is simply an aggregated version of the constraints $\langle \mathsf{C}_k, \mathsf{X}\rangle = 0$ which in turn follow from both $\langle \mathsf{A}_k, \mathsf{X}\rangle = 0$ and $\langle \mathsf{B}_k, \mathsf{X}\rangle = 0$. On the other hand, we already know (cf. the proof of Theorem 6.1) that $\langle \mathsf{M}_{m+1}, \mathsf{X}\rangle = 0$ imply $\mathsf{x} \in P$ for all $\mathsf{X} \in \mathcal{K}_\diamond \supset \mathcal{C}$, which means $\langle \mathsf{A}_k, \mathsf{X}\rangle = 0$ and, as argued above, also $\langle \mathsf{C}_k, \mathsf{X}\rangle = 0$, which entails $\langle \mathsf{B}_k, \mathsf{X}\rangle = \langle \mathsf{C}_k, \mathsf{X}\rangle - a_k\langle \mathsf{A}_k, \mathsf{X}\rangle = 0$ for all $k\in[1:p]$, i.e., X is (36)-feasible if it was (34)-feasible, and vice versa. Since above arguments hold also at the sub-zero level, all the primal equalities follow. On the dual side, we have, by a similar argument, for all $\bar{\mathsf{y}} = (\mathsf{y}, u_{m+1}) \in \mathbb{R} \times \mathbb{R}^m_+ \times \mathbb{R}$

$$\mathsf{Z}_{\mathrm{agg}}(\bar{\mathsf{y}}) = \mathsf{Z}(\mathsf{y}) + u_{m+1}\mathsf{M}_{m+1} = \mathsf{Z}_{\mathrm{Burer}}(\mathsf{y}, u_{m+1}\mathsf{a}, u_{m+1}\mathsf{e})\,,$$

which establishes $z^*_{CD,\mathrm{agg}} \le z^*_{CD,\mathrm{Burer}}$. Finally, if $z^*_{CD,\mathrm{agg}} = z^*_{CP,\mathrm{agg}}$ and likewise $z^*_{SD,\mathrm{agg}} = z^*_{SP,\mathrm{agg}}$, then (37) yields (38). □

A short summary of above results could be the following one: if Q_i is strictly copositive for at least one $i\in[0:m+1]$, then

$$z^*_{LD,+} \le z^*_{\rm semi} = z^*_{CP} \le z^*_{CP,\rm Burer} = z^*_{CP,\rm agg} \le z^*_+ \, .$$

As an aside, one may note that the two *inequality* constraints $\langle \mathsf{A}_k, \mathsf{X}\rangle \le 0$ and $\langle \mathsf{B}_k, \mathsf{X}\rangle \le 0$ already imply the *equalities* $\langle \mathsf{A}_k, \mathsf{X}\rangle = \langle \mathsf{B}_k, \mathsf{X}\rangle = 0$, whenever $\mathsf{X} \succeq \mathsf{O}$ with $\langle \mathsf{J}_0, \mathsf{X}\rangle = 1$ and all $a_k \ge 0$. Indeed, if $0 \le a_k \le \mathsf{r}_k^\top \mathsf{x}$, then squaring this inequality, using again the fact $\mathsf{Y} \succeq \mathsf{x}\mathsf{x}^\top$ and using $\langle \mathsf{B}_k, \mathsf{X}\rangle \le 0$ already entails

$$a_k^2 \le (\mathsf{r}_k^\top \mathsf{x})^2 \le \mathsf{r}_k^\top \mathsf{Y} \mathsf{r}_k \le a_k^2 \, ,$$

so $\langle \mathsf{A}_k, \mathsf{X}\rangle = \langle \mathsf{B}_k, \mathsf{X}\rangle = 0$ follows.

Interestingly, the idea to aggregate constraints in copositive optimization formulations recently emerged almost simultaneously and independently by the different approaches in [2, 23, 32]. However, very recent and preliminary empirical evidence on closely related problems [10] shows no clear advantage of either formulation, which is the reason why we mainly concentrated on the non-aggregated versions in this paper. See Section 7 for further discussion.

6.3 A further global optimality condition

As done in Corollary 5.1 in Subsection 5.2, we can also derive a second-order condition which guarantees global optimality of a generalized KKT point. Again, the slack matrix has to be copositive, and all we need is to adapt to the problem formulation with the redundant constraints à la Burer:

$$z^*_+ = \inf \{ q_0(\mathsf{x}) : \mathsf{x} \in F \cap P : q_i(\mathsf{x}) = 0 \, , \, i\in[m+1:p] \}$$

with $q_{m+k}(\mathsf{x}) = (\mathsf{r}_k^\top \mathsf{x})^2 - a_k^2 = 0$ as $k\in[1:p]$. In this context, a pair $(\mathsf{x}; \mathsf{u}, \mathsf{w}, \mathsf{z}) \in F \cap P \times \mathbb{R}^m_+ \times \mathbb{R}^{2p}$ is called *generalized KKT pair* if and only if

$$\left. \begin{array}{rcll} x_j[(\mathsf{H}_\mathsf{u} + \sum\limits_k z_k \mathsf{r}_k \mathsf{r}_k^\top)\mathsf{x} - \mathsf{d}_\mathsf{u} - \mathsf{A}^\top \mathsf{w}]_j &=& 0 & \text{for all } j\in[1:n] \text{ and} \\ u_i q_i(\mathsf{x}) &=& 0 & \text{for all } i\in[1:m] \, . \end{array} \right\} \tag{39}$$

Again (39) is equivalent to requiring that x is a critical point of the Lagrangian function, but without imposing sign constraints on the multipliers of the sign constraints $x_j \ge 0$.

Theorem 6.4 *If at a generalized KKT pair* $(\bar{\mathsf{x}}; \bar{\mathsf{u}}, \bar{\mathsf{w}}, \bar{\mathsf{z}}) \in F \cap P \times \mathbb{R}^m_+ \times \mathbb{R}^{2p}$ *in the sense of (39), the matrix*

$$\left[\begin{array}{cc} \mathsf{c}^\top \bar{\mathsf{u}} - \sum\limits_k \bar{z}_k a_k^2 + 2\mathsf{a}^\top \bar{\mathsf{w}} - q_0(\bar{\mathsf{x}}), & -(\mathsf{d}_{\bar{\mathsf{u}}} + \mathsf{A}^\top \bar{\mathsf{w}})^\top \\ -(\mathsf{d}_{\bar{\mathsf{u}}} + \mathsf{A}^\top \bar{\mathsf{w}}), & \mathsf{H}_{\bar{\mathsf{u}}} + \sum\limits_k \bar{z}_k \, \mathsf{r}_k \mathsf{r}_k^\top \end{array} \right] \tag{40}$$

is copositive, then $\bar{\mathsf{x}}$ *is a global solution to (2).*

Proof. The proof is similar to, but even simpler than, the proof of the implication $(c) \implies (d)$ in Theorem 5.1. In fact, condition (39) implies here $\bar{\mathsf{x}}^\top \mathsf{H}_{\bar{\mathsf{u}}} \bar{\mathsf{x}} = \bar{\mathsf{x}}^\top \mathsf{d}_{\bar{\mathsf{u}}}$, so that $\bar{\mathsf{X}} = \bar{\mathsf{z}}\bar{\mathsf{z}}^\top$ with $\bar{\mathsf{z}}^\top = [1, \bar{\mathsf{x}}^\top]$ forms an optimal primal-dual pair $(\bar{\mathsf{X}}; \bar{\mathsf{y}}, \bar{\mathsf{w}}, \bar{\mathsf{z}})$ to the copositive problem (36), if we define $\bar{\mathsf{y}}^\top = [q_0(\bar{\mathsf{x}}), \bar{\mathsf{u}}^\top]$, in which case the matrix in (40) exactly is $\mathsf{Z}_{\mathrm{Burer}}(\bar{\mathsf{y}}, \bar{\mathsf{w}}, \bar{\mathsf{z}})$. □

As before, specializing $\bar{\mathsf{w}} = \bar{u}_{m+1}\mathsf{a}$ and $\bar{\mathsf{z}} = \bar{u}_{m+1}\mathsf{e}$, the matrix in (40) simplifies to

$$\begin{bmatrix} \mathsf{c}^\top \bar{\mathsf{u}} + \bar{u}_{m+1}\|\mathsf{a}\|^2 - q_0(\bar{\mathsf{x}}), & -(\mathsf{d}_{\bar{\mathsf{u}}} + \bar{u}_{m+1}\mathsf{A}^\top \mathsf{a})^\top \\ -(\mathsf{d}_{\bar{\mathsf{u}}} + \bar{u}_{m+1}\mathsf{A}^\top \mathsf{a}), & \mathsf{H}_{\bar{\mathsf{u}}} + \bar{u}_{m+1}\mathsf{A}^\top \mathsf{A} \end{bmatrix},$$

which exactly corresponds to the (generalized) KKT formulation for $z_+ = \inf\left\{\mathsf{x} \in F \cap \mathbb{R}^n_+ : \|\mathsf{A}\mathsf{x} - \mathsf{a}\|^2 = 0\right\}$ with multiplier $\bar{u}_{m+1}$ for the last constraint.

7 Possible algorithmic implications

7.1 Update on approximation hierarchies

Both cones $\mathcal{C}$ and $\mathcal{C}^\star$ involved in the primal-dual pair (12) and (13) are intractable. So we need to approximate them by so-called *hierarchies,* i.e., a sequence of tractable cones $\mathcal{K}_d^\star$ such that $\mathcal{K}_d^\star \subset \mathcal{K}_{d+1}^\star \subset \mathcal{C}^\star$ where d is the level of the hierarchy, and $\bigcup_{d=0}^\infty \mathcal{K}_d^\star = [\mathcal{C}^\star]^\circ$, i.e., every strictly copositive matrix is contained in $\mathcal{K}_d^\star$ for some d. On the dual side, $\mathcal{K}_d$ are also tractable, $\mathcal{K}_{d+1} \subset \mathcal{K}_d$, and $\bigcap_{d=0}^\infty \mathcal{K}_d = \mathcal{C}$ contains no matrix which is not completely positive. For brevity of exposition, assume that $z^*_{CD} = z^*_{CP}$ and further assume that strong duality also holds for the approximation:

$$\begin{aligned} z^*_{\mathcal{K}_d} &:= \min\left\{\langle \mathsf{M}_0, \mathsf{X}\rangle : \langle \mathsf{M}_i, \mathsf{X}\rangle \le r_i\,,\ i \in [1\!:\!m]\,,\ \mathsf{X} \in \mathcal{K}_d\right\} \\ &= \max\left\{\mathsf{r}^\top \mathsf{y} : \mathsf{y} \in \mathbb{R}^m_+\,,\ \mathsf{M}_0 + \sum_{i=1}^m y_i \mathsf{M}_i \in \mathcal{K}_d^\star\right\}. \end{aligned}$$

Then by above we get $z^*_{\mathcal{K}_d} \to z^*_{CD} = z^*_{CP}$ as $d \to \infty$. By now, there are many possibilities explored for hierarchies $(\mathcal{K}_d)_d$, for a concise survey see [11]. Many of these involve linear or psd constraints of matrices of order n^{d+2}, e.g. the seminal ones proposed in [33, 40]. In particular for LMIs, matrices of larger order pose a serious memory problem for algorithmic implementations even for moderate d if n is large. LP-based hierarchies suffer less from this curse of dimensionality, and therefore we will follow a compromise between

LP-based and SDP-based hierarchies. We start with the usual zero-order approximation by the cone of *doubly nonnegative (DNN) matrices*

$$\mathcal{K}_0 = \{\mathsf{X} \text{ is psd} : \mathsf{X} \text{ has no negative entries}\} \ . \tag{41}$$

For the dual cone

$$\mathcal{K}_0^\star = \{\mathsf{P} + \mathsf{N} : \mathsf{P} \text{ is psd. and } \mathsf{N} \text{ has no negative entries}\} \tag{42}$$

Florian Jarre (personal communication) very recently has coined the term *nonnegative decomposable (NND)* for matrices in $\mathcal{K}_0^\star$, using the duality calculus pun $(DNN)^\star = NND$. Anyhow, based upon this construction, we may add valid linear inequalities, e.g., as done in [15, 16], yielding polyhedral inner approximations $\mathcal{L}_d^\star$ of the copositive cone, and, on the dual side, polyhedral outer approximations $\mathcal{L}_d$ for the completely positive cone, and finally define

$$\mathcal{K}_d := \mathcal{K}_0 \cap \mathcal{L}_d \ , \ d \in \{0, 1, 2, \dots\} \ , \tag{43}$$

or, by duality, the closure $\mathcal{K}_d^\star$ of the Minkowski sum $\mathcal{K}_0^\star + \mathcal{L}_d^\star$. Of course, this approximation satisfies above properties of exhaustivity, and involves LMIs only for matrices of order linear in n; in fact, we only employ the matrices $\mathsf{M}_i = \mathsf{M}(q_i)$ of order $n + 1$.

A similar yet different approach is taken in [34] where a conic *exact reformulation* of problem (1) is proposed, using another intractable cone, and constructing tractable approximation hierarchies for this cone. The examples specified in [34] reduce again to the NND cone $\mathcal{K}_0^\star$ or its dual, the DNN cone $\mathcal{K}_0$. However for large n, even $\mathcal{K}_0$ may involve too many (namely $\frac{(n-1)n}{2}$) linear inequalities to allow for efficient computation. This problem can be overcome by warmstarting as in [25], identifying or separating valid linear inequalities on the fly, or by the recently proposed tightening and acceleration method in [32].

The following proposal is an alternative: suppose that we only employ, say, n inequalities, e.g., by forbidding negative entries only in the first row of a matrix, to proxy for complete positivity. Then we arrive at $\mathcal{K}_\diamond = \{\mathsf{X} \text{ is psd} : X_{0j} \geq 0 \text{ for all } j \in [1:n]\}$ introduced in (18), and used in the SDP reformulation of the full Lagrangian dual in Subsection 3.2. Above discussion now justifies the term *sub-zero level approximation*.

A possibly efficient hierarchy is then

$$\mathcal{K}_{\diamond,d} = \mathcal{K}_\diamond \cap \mathcal{L}_d \ , \ d \in \{0, 1, 2, \dots\} \ . \tag{44}$$

While practical experience with this proposal is not yet available, we have seen above that $\mathcal{K}_{\diamond,d}$ emerges quite naturally in the context of Lagrangian duality and thus can be seen as a conceptual way of selecting (few) linear inequality constraints to tighten the SDP bound.

7.2 Approximate copositive bounds dominate Lagrangian dual bounds even at (sub-)zero level

Recall that the dual cone of $\mathcal{K}_\diamond$ is given by

$$\mathcal{K}_\diamond^\star = \left\{ \mathsf{P} + \begin{bmatrix} 0 & \mathsf{v}^\top \\ \mathsf{v} & \mathsf{O} \end{bmatrix} : \mathsf{P} \text{ is psd}\, , \, \mathsf{v} \in \mathbb{R}^n_+ \right\} . \tag{45}$$

The fact that every positive-semidefinite matrix lies in $\mathcal{K}_\diamond^\star$ is another reflection of the relation $z^*_{LD} \le z^*_{LD,+}$. On the other side, we by now can easily see that even at (sub-)zero level of approximation, the resulting tractable bound tightens the Lagrangian bound:

Theorem 7.1 *Consider any approximation hierarchy $\mathcal{K}_{\diamond,d}$ starting with $\mathcal{K}_\diamond$ as defined in (18), e.g. the one defined in (44), together with their bounds* $z^*_{\mathcal{K}_{\diamond,d}} = \inf\left\{ \langle \mathsf{M}_0, \mathsf{X}\rangle : \langle \mathsf{M}_i, \mathsf{X}\rangle \le r_i\, ,\ i{\in}[1{:}m]\, ,\ \mathsf{X} \in \mathcal{K}_{\diamond,d}\right\}$. *Then*

$$z^*_{LD,+} \le z^*_{\mathcal{K}_{\diamond,d}} \quad \textit{for all } d \in \{0,1,\ldots\}\ ,$$

and $z^*_{\mathcal{K}_{\diamond,d}} \uparrow z^*_{\text{semi}}$ as $d \to \infty$.

Proof. The inclusions $\mathcal{K}_\diamond^\star \subseteq \mathcal{K}_{\diamond,d}^\star$ and/or $\mathcal{K}_{\diamond,d} \subseteq \mathcal{K}_\diamond$ imply the inequality for all d, while exhaustivity $\bigcap_{d=0}^{\infty} \mathcal{K}_{\diamond,d} = \mathcal{C}$ yields $z^*_{\mathcal{K}_{\diamond,d}} \uparrow z^*_{\text{semi}}$ as $d \to \infty$. □

Example, continued from Section 5.2. Now assume for sake of illustration that for some d we have

$$\mathcal{K}_{\diamond,d} \subseteq \left\{ \mathsf{X} \in \mathcal{K}_\diamond : \sum_{i=1}^n X_{ii} \le X_{00} \right\} = (\mathbb{R}_+ \mathsf{D})^\star \quad \text{with} \quad \mathsf{D} = \begin{bmatrix} 1 & \mathsf{o}^\top \\ \mathsf{o} & -\mathsf{I}_n \end{bmatrix} .$$

Then for $y_0 = \lambda_{\min}(\mathsf{Q}_0) < 0$ we get $\mathsf{Z}_+(y_0, \mathsf{o}) + y_0 \mathsf{D} \succeq \mathsf{O}$ and therefore $\mathsf{Z}_+(y_0,\mathsf{o}) \in \mathcal{K}_d^\star$, because $-y_0\mathsf{D} \in \mathbb{R}_+\mathsf{D} \subseteq \mathcal{K}_{\diamond,d}^\star$, and because $\mathcal{K}_{\diamond,d}^\star \supseteq \mathcal{K}_\diamond^\star + \mathcal{L}_d^\star$ also includes all psd matrices by (45). We can conclude

$$-\infty = z^*_{LD,+} < y_0 \le z^*_{\mathcal{K}_{\diamond,d}}\, ,$$

so the gap is significantly reduced even by adding a single, very basic linear constraint to the starting cone $\mathcal{K}_\diamond$. Obviously, if an instance Q_0 is indefinite but satisfies $\mathsf{Z}_+(0,\mathsf{o}) = \begin{bmatrix} 0 & \mathsf{o}^\top \\ \mathsf{o} & \mathsf{Q}_0 \end{bmatrix} \in \mathcal{K}_{\diamond,d}^\star$, then we even have *closed the gap at finite level d* in the new hierarchy, while the Lagrangian duality gap is still infinite:

$$-\infty = z^*_{LD,+} < z^*_{\mathcal{K}_{\diamond,d}} = z^*_{CD} = z^*_{\text{semi}} = z^*_+ = 0\, .$$

Burer's relaxation simply adds another constraint to every linear equality constraint of the natural copositive formulation of the Semi-Lagrangian bound. Replacing $\mathcal{C}$ with $\mathcal{K}_d$ or $\mathcal{C}^\star$ with $\mathcal{K}_d^\star$, would therefore tighten the approximate bounds even beyond the Semi-Lagrangian dual, at the cost of dealing with additional constraints. As always in implementation, we have to face a trade-off between quality and effort of obtaining tractable bounds. Hopefully some empirical evidence will be put forward soon.

8 Conclusion and outlook

This paper deals with problems to optimize a quadratic function subject to quadratic and linear constraints, where the linear ones are treated separately. By relaxing everything except the sign constraints, we arrived at a Semi-Lagrangian dual which apparently has not been analyzed before in the literature. Here we have reformulated both the Lagrangian dual and the Semi-Lagrangian dual as conic optimization problems, and compared the resulting bounds to their counterparts when all linear equality constraints are replaced by a single convex quadratic one. This alternative turned out to be essentially equivalent to Burer's copositive relaxation. While the Semi-Lagrangian dual is a copositive problem, the Lagrangian dual can be seen as a natural relaxation of the latter, namely arising from an approximation of the copositive problem at a sub-zero level. This low level is important in regimes where every additional linear inequality constraint severely slows down algorithmic performance and/or creates memory problems, which is typical for interior-point methods when applied to very large problems, for instance in the most familiar doubly-nonnegative relaxation. For an interesting review of these and related bounds (as known prior to 2011), we refer to the survey article [4].

The development led us to propose a new variant building upon known approximation hierarchies which may avoid above drawbacks, with the hope that a significant tightening of the bounds becomes tractable, because LMIs of higher order matrices can be avoided. Furthermore, we studied properties of the problem which ensure strong duality of the conic relaxations; specified necessary and sufficient copositivity-based conditions to guarantee that the Semi-Lagrangian relaxation is exact; and proposed a hierarchy of seemingly new, sufficient, second-order global optimality conditions for a KKT point of the original problem which can be tested in polynomial time if tractable approximation hierarchies are employed. These conditions require much less than the familiar ones which require positive-semidefiniteness of the Hessian of the Lagrangian.

Building upon these findings, there are several directions of future research, among them:

- to tighten other variants of SDP formulations of the full Lagrangian relaxation [26], and to interpret them in terms of properties of the Lagrangian function of the original problem (in some formulation);
- to define a strategy which balances computational effort identifying and using additional linear constraints (i.e., other than those defining $\mathcal{K}_\diamond$), against efficient strengthening of the resulting bounds;
- to explore the quality of the relaxation if the A_k constraints are simply replaced by the B_k constraints, and to relate the result with above dual bounds.

Acknowledgement. The author is indebted to Associate Editor Samuel Burer and to three anonymous referees for their diligence and many suggestions which helped to significantly improve presentation of the paper. I am grateful for valuable comments and stimulating discussions on an earlier draft of this paper, provided by Peter J.C. Dickinson and Luis Zuluaga. Thanks are also due to the Isaac Newton Institute at Cambridge University for providing a stimulating environment when the author participated as a visiting fellow in the Polynomial Optimization Programme 2013, organized by Joerg Fliege, Jean Bernard Lasserre, Adam Letchford and Markus Schweighofer.

References

[1] Wenbo Ai and Shuzhong Zhang. Strong duality for the CDT subproblem: a necessary and sufficient condition. *SIAM J. Optim.*, 19(4):1735–1756, 2009.

[2] Naohiko Arima, Sunyoung Kim, and Masakazu Kojima. Simplified copositive and Lagrangian relaxations for linearly constrained quadratic optimization problems in continuous and binary variables. *Pacific J. Optimiz.*, 10(3):437–451, 2014.

[3] Lijie Bai, John E.Mitchell, and Jong-Shi Pang. On QPCCs, QCQPs and completely positive programs. Preprint, RPI, `http://www.optimization-online.org/DB\_HTML/2014/01/4213.html`, 2014.

[4] Xiaowei Bao, Nikolaos V. Sahinidis, and Mohit Tawarmalani. Semidefinite relaxations for quadratically constrained quadratic programming: a review and comparisons. *Math. Program.*, 129(Ser. B):129–157, 2011.

[5] Amir Beck and Yonina Eldar. Strong duality in nonconvex quadratic optimization with two quadratic constraints. *SIAM J. Optim.*, 17(3):844–860, 2006.

[6] Evgeny G. Belousov and Diethard Klatte. A Frank-Wolfe type theorem for convex polynomial programs. *Comput. Optim. Appl.*, 22(1):37–48, 2002.

[7] Aharon Ben Tal, Arkadi Nemirovski, and Cornelis Roos. Robust solutions of uncertain quadratic and conic-quadratic problems. *SIAM J. Optim.*, 13(2):535–560, 2002.

[8] Immanuel M. Bomze. Copositive optimization – recent developments and applications. *European J. Oper. Res.*, 216:509–520, 2012.

[9] Immanuel M. Bomze. Copositivity for second-order optimality conditions in general smooth optimization problems. Preprint NI13033-POP, Isaac Newton Institute, Cambridge UK, 2013.

[10] Immanuel M. Bomze, Jianqiang Cheng, Peter J.C. Dickinson, and Abdel Lisser. A fresh CP look at mixed-binary QPs: new formulations, relaxations and penalizations. Preprint, LRI, Univ. Paris Sud, 2015 (in preparation).

[11] Immanuel M. Bomze, Mirjam Dür, and Chung-Piaw Teo. Copositive optimization. *Optima – MOS Newsletter*, 89:2–10, 2012.

[12] Immanuel M. Bomze and Florian Jarre. A note on Burer's copositive representation of mixed-binary QPs. *Optim. Lett.*, 4(3):465–472, 2010.

[13] Immanuel M. Bomze and Michael L. Overton. Narrowing the difficulty gap for the Celis-Dennis-Tapia problem. *Math. Program.*, to appear. DOI:10.1007/s10107–014–0836–3, 2015.

[14] Immanuel M. Bomze, Werner Schachinger, and Reinhard Ullrich. New lower bounds and asymptotics for the cp-rank. *SIAM J. Matrix Anal. Appl.*, 36(1):20–37, 2015.

[15] Stefan Bundfuss and Mirjam Dür. Algorithmic copositivity detection by simplicial partition. *Linear Algebra Appl.*, 428(7):1511–1523, 2008.

[16] Stefan Bundfuss and Mirjam Dür. An adaptive linear approximation algorithm for copositive programs. *SIAM J. Optim.*, 20(1):30–53, 2009.

[17] Samuel Burer. On the copositive representation of binary and continuous nonconvex quadratic programs. *Math. Program.*, 120(2, Ser. A):479–495, 2009.

[18] Samuel Burer. Copositive programming. In Miguel F. Anjos and Jean Bernard Lasserre, editors, *Handbook of Semidefinite, Cone and Polynomial Optimization: Theory, Algorithms, Software and Applications*, International Series in Operations Research and Management Science, pages 201–218. Springer, New York, 2012.

[19] Samuel Burer, Sunyoung Kim, and Masakazu Kojima. Faster, but weaker, relaxations for quadratically constrained quadratic programs. *Comput. Optim. Appl.*, 59(1–2):27–45, 2014.

[20] Jieqiu Chen and Samuel Burer. Globally solving nonconvex quadratic programming problems via completely positive programming. *Math. Program. Comput.*, 4(1):33–52, 2012.

[21] Peter J. C. Dickinson. An improved characterisation of the interior of the completely positive cone. *Electron. J. Linear Algebra*, 20:723–729, 2010.

[22] Peter J. C. Dickinson and Luuk Gijben. On the computational complexity of membership problems for the completely positive cone and its dual. *Comput. Optim. Appl.*, 57(2):403–415, 2014.

[23] Peter J.C. Dickinson. *The copositive cone, the completely positive cone and their generalisations.* Ph.D. dissertation, University of Groningen, The Netherlands, 2013.

[24] Mirjam Dür. Copositive programming — a survey. In Moritz Diehl, Francois Glineur, Elias Jarlebring, and Wim Michiels, editors, *Recent Advances in Optimization and its Applications in Engineering*, pages 3–20. Springer, Berlin Heidelberg New York, 2010.

[25] Alexander Engau, Miguel F. Anjos, and Immanuel M. Bomze. Constraint selection in a build-up interior-point cutting-plane method for solving relaxations of the stable-set problem. *Math. Methods of OR*, 78(1):35–59, 2013.

[26] Alain Faye and Frédéric Roupin. Partial Lagrangian relaxation for general quadratic programming. *4OR*, 5:75–88, 2007.

[27] Marguerite Frank and Philip Wolfe. An algorithm for quadratic programming. *Naval Res. Logistics Quarterly*, 3:95–110, 1956.

[28] Tetsuya Fujie and Masakazu Kojima. Semidefinite programming relaxation for nonconvex quadratic programs. *J. Global Optim.*, 10(4):367–380, 1997.

[29] Simai He, Zhi-Quan Luo, Jiawang Nie, and Shuzhong Zhang. Semidefinite relaxation bounds for indefinite homogeneous quadratic optimization. *SIAM J. Optim.*, 19(2):503–523, 2008.

[30] Vaithilingam Jeyakumar, Gue Myung Lee, and Guoyin Li. Alternative theorems for quadratic inequality systems and global quadratic optimization. *SIAM J. Optim.*, 20(2):983–1001, 2009.

[31] Vaithilingam Jeyakumar and Guoyin Li. Trust-region problems with linear inequality constraints: exact SDP relaxation, global optimality and robust optimization. *Math. Program.*, 147(Ser. A):171–206, 2014.

[32] Sunyoung Kim, Masakazu Kojima, and Kim-Chuan Toh. A Lagrangian-DNN relaxation: a fast method for computing tight lower bounds for a class of quadratic optimization problems. *Math. Program.*, to appear. DOI:10.1007/s10107–015–0874–5, 2015.

[33] Jean Bernard Lasserre. Global optimization with polynomials and the problem of moments. *SIAM J. Optim.*, 11(3):796–817, 2000/01.

[34] Cheng Lu, Shu-Cherng Fang, Qingwei Jin, Zhengbo Wang, and Wenxun Xing. KKT solution and conic relaxation for solving quadratically constrained quadratic programming problems. *SIAM J. Optim.*, 21(4):1475–1490, 2011.

[35] Zhi-Quan Luo and Shuzhong Zhang. On extensions of the Frank-Wolfe theorems. *Comput. Optim. Appl.*, 13(1-3):87–110, 1999.

[36] Jérome Malick and Frédéric Roupin. On the bridge between combinatorial optimization and nonlinear optimization: a family of semidefinite bounds for 0-1 quadratic problems leading to quasi-Newton methods. *Math. Program.*, 140(Ser. B):99–124, 2013.

[37] Katta G. Murty and Santosh N. Kabadi. Some NP-complete problems in quadratic and nonlinear programming. *Math. Program.*, 39(2):117–129, 1987.

[38] Yuri Nesterov, Henry Wolkowicz, and Yinyu Ye. Nonconvex quadratic optimization. In Henry Wolkowicz, Romesh Saigal, and Lieven Vandenberghe, editors, *Handbook of Semidefinite Programming*, pages 361–420. Kluwer, New York, 2000.

[39] Arnold Neumaier. Optimization – Theory and Algorithms. Preprint, `http://www.mat.univie.ac.at/~neum/`, 2014.

[40] Pablo A. Parrilo. *Structured Semidefinite Programs and Semi-algebraic Geometry Methods in Robustness and Optimization.* PhD thesis, California Institute of Technology, Pasadena, CA, May 2000.

[41] Svatopluk Poljak, Franz Rendl, and Henry Wolkowicz. A recipe for semidefinite relaxation for (0,1)-quadratic programming. *J. Global Optim.*, 7(1):51–73, 1995.

[42] Motakuri Venkata Ramana. *An Algorithmic Analysis of Multiquadratic and Semidefinite Programming Problems.* Doctoral dissertation, Johns Hopkins University, Baltimore, MD, 1993.

[43] Naomi Shaked-Monderer, Abraham Berman, Immanuel M. Bomze, Florian Jarre, and Werner Schachinger. New results on the cp rank and related properties of co(mpletely)positive matrices. *Linear Multilinear Algebra*, 63(2):384–396, 2015.

[44] Naum Z. Shor. Quadratic optimization problems. *Izv. Akad. Nauk SSSR Tekhn. Kibernet.*, 222(1):128–139, 1987.

[45] Yinyu Ye. On the complexity of approximating a KKT point of quadratic programming. *Math. Program.*, 80(Ser. A):195–211, 1998.