Copositivity for second-order optimality conditions in general smooth optimization problems

Immanuel M. Bomze

ISOR, University of Vienna, Austria

Abstract

Second-order local optimality conditions involving copositivity of the Hessian of the Lagrangian on the reduced (polyhedral) tangent cone have the advantage that there is only a small gap between sufficient (the Hessian is strictly copositive) and necessary (the Hessian is copositive) conditions. In this respect, this is a proper generalization of convexity of the Lagrangian. We also specify a copositivity-based variant which is sufficient for global optimality. For (nonconvex) quadratic optimization problems over polyhedra (QPs), the distinction between sufficiency and necessity vanishes, both for local and global optimality. However, in the strictly copositive case we can provide a distance lower (error) bound of the increment $f(\mathsf{x})-f(\bar{\mathsf{x}})$ around a local minimizer $\bar{\mathsf{x}}$. This is a refinement of an earlier result which focussed on mere (non-strict) copositivity. In addition, an apparently new variant of constraint qualification (CQ) is presented which is implied by Abadie's CQ and which is suitable for second-order analysis. This new *reflected Abadie CQ* is neither implied, nor implies, Guignard's CQ. However, it implies the necessary second-order local optimality condition based on copositivity. Applications to trust-region and all-quadratic problems illustrate the advantage of this approach, by applying above proof techniques and several (counter-)examples.

Key words: Copositive matrices, non-convex optimization, global optimality condition, polynomial optimization, trust region problem, all-quadratic problem

November 20, 2013

1 Second-order local optimality conditions for general smooth nonlinear optimization

1.1 Constraint qualifications

To begin with, let us shortly recapitulate several *constraint qualifications (CQ)* for a smooth optimization problem

$$\begin{array}{rcll} f(\mathsf{x}) & \to & \min\ ! & \text{subject to} \\ h_i(\mathsf{x}) & = & 0\,, & i \in \{1,\dots,q\}\ , \\ g_i(\mathsf{x}) & \leq & 0\,, & i \in \{1,\dots,m\}\ , \end{array}$$

which can be written in the more compact form

$$\min_{\mathsf{x}\in M} f(\mathsf{x}) \quad \text{with} \quad M = \left\{\mathsf{x} \in \mathbb{R}^n : H(\mathsf{x}) = \mathsf{o} \text{ and } -G(\mathsf{x}) \in \mathbb{R}^m_+\right\}\ , \qquad (1)$$

$H(\mathsf{x}) = [h_1(\mathsf{x}), \dots h_q(\mathsf{x})]^\top \in \mathbb{R}^q$ and $G(\mathsf{x}) = [g_1(\mathsf{x}), \dots, g_m(\mathsf{x})]^\top \in \mathbb{R}^m$. Note that the following conditions depend on f and the current description of M by G and H, not only on the shape of M and f. All functions f, G and H are supposed to have continuous second-order derivatives (the derivatives w.r.t. x are symbolized by D_x, sometimes also by $\nabla f = [D_\mathsf{x} f]^\top$, while $\dot\phi(t) = \frac{d\phi}{dt}$ denotes derivative w.r.t. scalar variable t). As usual, for any cone $C \subseteq \mathbb{R}^n$, we denote its dual cone by

$$C^\star = \left\{\mathsf{u} \in \mathbb{R}^n : \mathsf{u}^\top \mathsf{v} \geq 0 \text{ for all } \mathsf{v} \in C\right\}\ .$$

Definition 1.1 *Let $\mathsf{x} \in M$ be a feasible point of problem (1) and denote by $I(\mathsf{x}) = \{i \in \{1,\dots,m\} : g_i(\mathsf{x}) = 0\}$ the indices of constraints binding at x, as well as by*

$$\Gamma(\mathsf{x}) = \left\{\mathsf{v} \in \mathbb{R}^n : D_\mathsf{x} H(\mathsf{x})\mathsf{v} = \mathsf{o} \textit{ and } \mathsf{v}^\top \nabla g_i(\mathsf{x}) \leq 0 \textit{ for all } i \in I(\mathsf{x})\right\}$$

the polyhedral tangent cone *of M at x. Later on, we will also use the* reduced polyhedral tangent cone

$$\Gamma_0(\mathsf{x}) = \left\{\mathsf{v} \in \Gamma(\mathsf{x}) : \mathsf{v}^\top \nabla f(\mathsf{x}) = 0\right\} = \Gamma(\mathsf{x}) \cap \nabla f(\mathsf{x})^\perp\,.$$

Finally, consider the (derivative) tangent cone

$$T_M(\mathsf{x}) = \left\{\mathsf{v} \in \mathbb{R}^n : \mathsf{v} = \lim_{s\searrow 0} \mathsf{v}_s \textit{ with } \mathsf{x} + t_s\mathsf{v}_s \in M,\ \textit{some } t_s \searrow 0 \textit{ as } s \searrow 0\right\}\ .$$

Note that, by taking directional derivatives, we always have $T_M(\mathsf{x}) \subseteq \Gamma(\mathsf{x})$ but the latter cone may be larger in general (and the former may neither be convex nor closed, let alone polyhedral).

(a) *We say that* G, H *satisfy the* linear independence CQ (LICQ) *at* x *if the gradients of binding constraints*

$$[\nabla h_i(\mathsf{x}) : i \in \{1, \ldots, q\}\,; \nabla g_i(\mathsf{x}) : i \in I(\mathsf{x})]$$

are linearly independent;

(b) *we say that* G, H *satisfy the* Mangasarian/Fromovitz CQ (MFCQ) *at* x *if the gradients*

$$[\nabla h_i(\mathsf{x}) : i \in \{1, \ldots, q\}]$$

are linearly independent (i.e., if rank $D_{\mathsf{x}} H(\mathsf{x}) = q$*) and if there is a direction* $\mathsf{d} \in \Gamma(\mathsf{x})$ *satisfying*

$$\mathsf{d}^\top \nabla g_i(\mathsf{x}) < 0 \quad \textit{for all } i \in I(\mathsf{x})\,;$$

(c) *we say that* G, H *satisfy the* Abadie CQ (ACQ) *at* x *if* $T_M(\mathsf{x}) = \Gamma(\mathsf{x})$*;*

(d) *we say that* G, H *satisfy the* Guignard CQ (GCQ) *at* x *if* $[T_M(\mathsf{x})]^\star = [\Gamma(\mathsf{x})]^\star$*;*

(e) *finally, we define an apparently new CQ which we propose to call* reflected ACQ (RACQ): *we say that* G, H *satisfy the RACQ at* x *if*

$$\Gamma(\mathsf{x}) \subseteq T_M(\mathsf{x}) \cup [-T_M(\mathsf{x})]\,;$$

in sloppy words, RACQ are satisfied if and only if for any direction $\mathsf{v} \in \Gamma(\mathsf{x})$*, either* v *or* $-\mathsf{v}$ *is a starting tangent of a trajectory (or curve) starting at* x *and remaining entirely inside* M*.*

For the readers' convenience, we detail the only non-trivial relation of above CQs; see, e.g. [6, Cor.12.1].

Lemma 1.1 *Suppose that* $\mathsf{v} \in \Gamma(\bar{\mathsf{x}})$ *satisfies* $\mathsf{v}^\top \nabla g_i(\bar{\mathsf{x}}) < 0$ *for all* $i \in I(\bar{\mathsf{x}})$*, and further suppose that rank* $D_{\mathsf{x}} H(\bar{\mathsf{x}}) = q$*. Then there is a trajectory* $\mathsf{y}(t) \in \mathbb{R}^n$ *with* $\mathsf{y}(t) \in M$ *for all small enough* $t \geq 0$ *with* $\mathsf{y}(0) = \bar{\mathsf{x}}$*, and having a tangent* $\dot{\mathsf{y}}(0) = \mathsf{v}$*. Hence* $\mathsf{v} = \lim_{s \searrow 0} \mathsf{v}_s$ *with* $\mathsf{v}_s = \frac{1}{s}[\mathsf{y}(s) - \mathsf{y}(0)]$ *and* $\bar{\mathsf{x}} + s\mathsf{v}_s = \mathsf{y}(s) \in M$*.*

Proof. The $q \times n$ Jacobian matrix $D_{\mathsf{x}} H(\mathsf{x})$ has rows $[\nabla h_i(\mathsf{x})]^\top$, $i \in \{1, \ldots, q\}$. Now, for $\mathsf{w} \in \mathbb{R}^q$ and $t \in \mathbb{R}$, define the mapping $\Phi : \mathbb{R}^q \times \mathbb{R} \to \mathbb{R}^q$ by

$$\Phi(\mathsf{w}, t) := H(\bar{\mathsf{x}} + t\mathsf{v} + D_{\mathsf{x}} H(\bar{\mathsf{x}})^\top \mathsf{w})\,.$$

Then $\Phi(\mathsf{o}, 0) = H(\bar{\mathsf{x}}) = \mathsf{o}$ and, by assumption, the Jacobian $D_{\mathsf{w}}\Phi(\mathsf{o}, 0) = D_{\mathsf{x}} H(\bar{\mathsf{x}})[D_{\mathsf{x}} H(\bar{\mathsf{x}})]^\top$ is nonsingular as rank $D_{\mathsf{x}} H(\mathsf{x}) = q$. Thus the Implicit

Function Theorem guarantees existence of a differentiable trajectory $\mathsf{w} : (-\varepsilon, \varepsilon) \to \mathbb{R}^q$ with $\mathsf{w}(0) = \mathsf{o}$ and $\Phi(\mathsf{w}(t), t) = 0$ if $|t| < \varepsilon$. Further, we have $\dot{\mathsf{w}}(0) = -[D_{\mathsf{w}}\Phi(\mathsf{o}, 0)]^{-1} D_t\Phi(\mathsf{o}, 0)$. Now $D_t\Phi(\mathsf{o}, 0) = D_{\mathsf{x}} H(\bar{\mathsf{x}})\mathsf{v} = \mathsf{o}$ as $\mathsf{v} \in \Gamma(\bar{\mathsf{x}})$ implies $\mathsf{v} \perp \nabla h_i(\bar{\mathsf{x}})$ for all $i = 1, \ldots, q$. So also $\dot{\mathsf{w}}(0) = \mathsf{o}$. Now define the trajectory $\mathsf{y}(t) = \bar{\mathsf{x}} + t\mathsf{v} + \mathsf{w}(t)$ which satisfies $\dot{\mathsf{y}}(0) = \mathsf{v}$. Further, by construction, $H(\mathsf{y}(t)) = \Phi(\mathsf{w}(t), t) = \mathsf{o}$ whenever $|t| < \varepsilon$. For these t, if ε is small enough, we can enforce $g_i(\mathsf{y}(t)) < 0$ whenever $g_i(\bar{\mathsf{x}}) < 0$ by continuity. Finally, we get, possibly further reducing ε if necessary,

$$g_i(\mathsf{y}(t)) = g_i(\bar{\mathsf{x}}) + t\mathsf{v}^\top \nabla g_i(\bar{\mathsf{x}}) + \mathfrak{o}(t) < g_i(\bar{\mathsf{x}}) = 0 \text{ for all } i \in I(\bar{\mathsf{x}}) \text{ if } 0 < t < \varepsilon\,.$$

We conclude $\mathsf{y}(t) \in M$ whenever $0 \leq t < \varepsilon$, as desired. □

Most of the following relations between the CQs are well known; for a standard reference see [10].

Corollary 1.1 *The LICQ imply the MFCQ which in turn imply the ACQ which in turn imply both the RACQ and the GCQ.*

Proof. If all gradients of binding constraints at $\bar{\mathsf{x}}$ are linearly independent, the linear system in d

$$\begin{aligned} \mathsf{d}^\top \nabla h_i(\bar{\mathsf{x}}) &= 0\,, \quad i \in \{1, \ldots, q\}\,, \\ \mathsf{d}^\top \nabla g_i(\bar{\mathsf{x}}) &= -1\,, \quad i \in I(\bar{\mathsf{x}})\,, \end{aligned}$$

has a solution $\mathsf{d} \in \mathbb{R}^n$. Any such d must lie in $\Gamma(\bar{\mathsf{x}})$ and hence MFCQ holds. To show the remaining assertion, take any $\mathsf{v} \in \Gamma(\bar{\mathsf{x}})$, choose any small $s > 0$, and consider a direction $\mathsf{d} \in \Gamma(\bar{\mathsf{x}})$ which satisfies $\mathsf{d}^\top \nabla g_i(\bar{\mathsf{x}}) < 0$ for all i with $g_i(\bar{\mathsf{x}}) = 0$. This d exists by the MFCQ. Clearly, also $\mathsf{v}_s = \mathsf{v} + s\mathsf{d} \in \Gamma(\bar{\mathsf{x}})$ satisfies the assumption of Lemma 1.1, so for any small $s > 0$ there is a trajectory $\mathsf{y}_s(t) \in M$ starting at $\bar{\mathsf{x}}$, i.e. $\mathsf{y}_s(0) = \bar{\mathsf{x}}$, and having starting tangent $\dot{\mathsf{y}}_s(0) = \mathsf{v}_s$. Hence the ACQ are met. Obviously, ACQ implies both RACQ and GCQ. □

Remark 1.1 *This example is taken from James V. Burke's extremely helpful site http://www.math.washington.edu/˜burke/crs/408/. Consider* $n = 2$, $f(\mathsf{x}) = \mathsf{x}^\top \mathsf{x}$, $G(\mathsf{x}) = -\mathsf{x}$ *and* $H(\mathsf{x}) = x_1 x_2$. *The (only) global solution is* $\mathsf{x}^* = \mathsf{o}$, *where* $T_M(\mathsf{x}^*) = \left\{\mathsf{x} \in \mathbb{R}^2_+ : x_1 x_2 = 0\right\} \subset \Gamma(\mathsf{x}^*) = \mathbb{R}^2_+$, *so both ACQ and RACQ are violated while GCQ is satisfied as both dual cones are equal to* $\mathbb{R}^2_+$.

Remark 1.2 *Given any set of multipliers* $\mathsf{u} = [u_1, \ldots, u_m]^\top \in \mathbb{R}^m_+$ *for the inequality constraints* $g_i(\mathsf{x}) \le 0$*, we will employ the RACQ for a subproblem*

$$M_{\mathsf{u}} = \{\mathsf{y} \in M : g_i(\mathsf{y}) = 0 \text{ if } u_i > 0\} \ ,$$

i.e., for the objective function f*, the equality constraints* h_i *and* $g_i (u_i > 0)$*, and the inequality constraints for* $g_i(u_i = 0)$*. For a KKT point* $\bar{\mathsf{x}}$ *with multipliers* u*, the reduced tangent cone* $\Gamma_0(\bar{\mathsf{x}})$ *obviously coincides with the polyhedral tangent cone of* M_{u} *because* $\mathsf{v} \perp \nabla f(\bar{\mathsf{x}}) = -\sum_i u_i \nabla g_i(\bar{\mathsf{x}})$ *and* $\mathsf{v} \in \Gamma(\bar{\mathsf{x}})$ *implies* $\mathsf{v} \perp \nabla g_i(\bar{\mathsf{x}})$ *if* $u_i > 0$*. While the LICQ are inherited from* $[M, \Gamma(\bar{\mathsf{x}})]$ *by* $[M_{\mathsf{u}}, \Gamma_0(\bar{\mathsf{x}})]$*, neither MFCQ nor ACQ nor RACQ are inherited, as the following example [5, Ex.2.3] shows:*

Let $m = n = 2$ *and* $g_1(x_1, x_2) = e^{-x_1} + x_1 - x_2 - 1$ *as well as* $g_2(x_1, x_2) = e^{x_1} - x_1 - x_2 - 1$*, which have at* $\bar{\mathsf{x}} = \mathsf{o}$ *the same gradients* $\nabla g_1(\mathsf{o}) = \nabla g_2(\mathsf{o}) = [0, -1]^\top$*. Hence for*

$$M = \{\mathsf{x} \in \mathbb{R}^2 : g_i(\mathsf{x}) \le 0 \, , \, 1 \le i \le 2\}$$

even the MFCQ at the point $\bar{\mathsf{x}} = \mathsf{o}$ *are satisfied: indeed for* $\mathsf{v} = [0, 1]^\top$ *we obtain* $\mathsf{v}^\top \nabla g_i(\bar{\mathsf{x}}) < 0$ *for all* i*. The tangent cone is* $\Gamma(\bar{\mathsf{x}}) = \{\mathsf{v} \in \mathbb{R}^2 : v_2 \ge 0\}$*. For any objective function* f *with gradient* $\nabla f(\mathsf{o}) = [0, 1]^\top$*, the point* $\bar{\mathsf{x}} = \mathsf{o}$ *satisfies the KKT-conditions, any admissible set* u *of Lagrange multipliers fulfilling* $u_1 + u_2 = 1$*. Now*

$$\Gamma_0(\bar{\mathsf{x}}) = \{\mathsf{v} \in \mathbb{R}^2 : v_2 = 0\} \, .$$

If both $u_1 > 0$ *and* $u_2 > 0$*, then* $M_{\mathsf{u}} = \{\mathsf{o}\}$ *and ACQ, and likewise RACQ, is obviously violated. If, however,* $u_1 = 1$ *and* $u_2 = 0$*, then*

$$\begin{aligned} M_{\mathsf{u}} &= \{\mathsf{x} \in M : g_1(\mathsf{x}) = 0\} \\ &= \{\mathsf{x} \in \mathbb{R}^2 : x_2 \ge e^{x_1} - x_1 - 1 \text{ and } x_2 = e^{-x_1} + x_1 - 1\} \\ &= \{\mathsf{x} \in \mathbb{R}^2 : \sinh x_1 \ge x_1 \text{ and } x_2 = e^{-x_1} + x_1 - 1\} \\ &= \{\mathsf{x} \in \mathbb{R}^2 : x_1 \ge 0 \text{ and } x_2 = e^{-x_1} + x_1 - 1\} \, , \end{aligned}$$

which also violates ACQ, because $\mathsf{v} = [-1, 0]^\top \in \Gamma_0(\bar{\mathsf{x}})$ *cannot be a starting tangent vector of any trajectory in* M_{u} *starting in* $\bar{\mathsf{x}} = \mathsf{o}$*. But* $-\mathsf{v}$ *is such a tangent vector, so RACQ holds. Similarly, also for* $\mathsf{u} = [0, 1]^\top$*, ACQ is not met by*

$$M_{\mathsf{u}} = \{\mathsf{x} \in \mathbb{R}^2 : x_2 = e^{x_1} - x_1 - 1 \text{ and } x_1 \le 0\} \, .$$

Here $\mathsf{v} = [1, 0]^\top$ *is no starting tangent vector but* $-\mathsf{v}$ *is one, so again RACQ holds. All these examples also violate the GCQ.*

1.2 Second-order conditions for local optimality

For problem (1), we define the *Lagrangean function*

$$L(\mathsf{x};\mathsf{u}) = f(\mathsf{x}) + \sum_{i=1}^{m} u_i g_i(\mathsf{x}) + \sum_{i=1}^{q} u_{i+m} h_i(\mathsf{x}) \,,$$

where $u_i \geq 0$ for all $i \in \{1, \dots, m\}$ and $u_i \in \mathbb{R}$ for all $i \in \{m+1, \dots m+q\}$ are the Lagrange multipliers of the constraints.

We now are ready to prove necessary and sufficient second-order optimality conditions with only a small gap in-between them. A precursor using ACQ instead of RACQ, and applied to a more general setting, can be found in [3] who apparently had the final word up to now in a series of publications dealing with similar second-order optimality conditions (e.g., [4, 8]).

The key notion for formulating these conditions is that of *copositivity.* Given a symmetric $n \times n$ matrix Q and a cone $\Gamma \subseteq \mathbb{R}^n$, we say that

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

Strict copositivity generalizes positive-definiteness (all eigenvalues strictly positive) and copositivity generalizes positive-semidefiniteness (no eigenvalue strictly negative) of a symmetric matrix.

Theorem 1.1 *Let $\bar{\mathsf{x}}$ be a KKT point with Lagrange multipliers $\bar{\mathsf{u}}$.*

(a) If $D^2_{\mathsf{x}} L(\bar{\mathsf{x}};\bar{\mathsf{u}})$ is strictly $\Gamma_0(\bar{\mathsf{x}})$-copositive, then $\bar{\mathsf{x}}$ is a strict local minimizer of f over M. More precisely, there are $\varepsilon > 0$ and $\rho > 0$ such that

$$f(\mathsf{x}) \geq f(\bar{\mathsf{x}}) + \rho \|\mathsf{x} - \bar{\mathsf{x}}\|^2 \quad \textit{for all } \mathsf{x} \in M \ \textit{with } \|\mathsf{x}\| < \varepsilon \,.$$

(b) If $M_{\bar{\mathsf{u}}}$ satisfies RACQ at $\bar{\mathsf{x}}$ (in the sense of Remark 1.2), and if $\bar{\mathsf{x}}$ is a local minimizer of f over M, then $D^2_{\mathsf{x}} L(\bar{\mathsf{x}};\bar{\mathsf{u}})$ is $\Gamma_0(\bar{\mathsf{x}})$-copositive.

Proof. (a) Assume the contrary, so that there are $\mathsf{x}_s \in M$ with $t_s = \|\mathsf{x}_s - \bar{\mathsf{x}}\| \searrow 0$ and $\rho_s \to 0$ as $s \searrow 0$ with

$$f(\mathsf{x}_s) < f(\bar{\mathsf{x}}) + \rho_s t_s^2 \quad \text{as } s \searrow 0 \,. \tag{2}$$

Consider the directions $\mathsf{v}_s = \frac{1}{t_s}(\mathsf{x}_s - \bar{\mathsf{x}})$ of unit length and assume without loss of generality that $\mathsf{v}_s \to \mathsf{v}$ as $s \searrow 0$. Then $\mathsf{v} \in \Gamma(\bar{\mathsf{x}})$ as noted in Definition 1.1. Obviously

$$\mathsf{v}^\top \nabla f(\bar{\mathsf{x}}) = \lim_{s \searrow 0} \mathsf{v}_s^\top \nabla f(\bar{\mathsf{x}}) = \lim_{s \searrow 0} \frac{1}{t_s}[f(\mathsf{x}_s) - f(\bar{\mathsf{x}})] \leq \lim_{s \searrow 0} \frac{\rho_s t_s^2}{t_s} = 0 \,,$$

so that $\mathsf{v} \in \Gamma_0(\bar{\mathsf{x}}) \setminus \{\mathsf{o}\}$ and therefore by assumption $\mathsf{v}^\top D^2_{\mathsf{x}} L(\bar{\mathsf{x}}; \bar{\mathsf{u}})\mathsf{v} > 0$. Now we estimate, by help of (2),

$$\begin{aligned} f(\bar{\mathsf{x}}) + \rho_s t_s^2 &> f(\mathsf{x}_s) \geq L(\mathsf{x}_s; \bar{\mathsf{u}}) = L(\bar{\mathsf{x}}; \bar{\mathsf{u}}) + \tfrac{t_s^2}{2}\, \mathsf{v}_s^\top D^2_{\mathsf{x}} L(\bar{\mathsf{x}}; \bar{\mathsf{u}})\mathsf{v}_s + \mathfrak{o}(t_s^2) \\ &= f(\bar{\mathsf{x}}) + \tfrac{t_s^2}{2}\, \mathsf{v}_s^\top D^2_{\mathsf{x}} L(\bar{\mathsf{x}}; \bar{\mathsf{u}})\mathsf{v}_s + \mathfrak{o}(t_s^2)\,, \end{aligned}$$

subtract $f(\bar{\mathsf{x}})$ and divide by $t_s^2 > 0$, to arrive at

$$\rho_s \geq \tfrac{1}{2}\, \mathsf{v}_s^\top D^2_{\mathsf{x}} L(\bar{\mathsf{x}}; \bar{\mathsf{u}})\mathsf{v}_s + \mathfrak{o}(1) \geq \tfrac{1}{3}\, \mathsf{v}^\top D^2_{\mathsf{x}} L(\bar{\mathsf{x}}; \bar{\mathsf{u}})\mathsf{v} > 0$$

for all small enough $s > 0$, a contradiction.
(b) Suppose $\mathsf{v} \in \Gamma_0(\bar{\mathsf{x}})$ satisfies $\mathsf{v}^\top D^2_{\mathsf{x}} L(\bar{\mathsf{x}}; \bar{\mathsf{u}})\mathsf{v} < 0$. Further, assume without loss of generality that $\mathsf{v} \in T_{M_{\bar{\mathsf{u}}}}(\bar{\mathsf{x}})$; indeed, otherwise replace v with $-\mathsf{v}$ which won't change the quadratic form $\mathsf{v}^\top D^2_{\mathsf{x}} L(\bar{\mathsf{x}}; \bar{\mathsf{u}})\mathsf{v}$. Then choose a close enough direction v_s and step sizes $t_s \searrow 0$ as $s \searrow 0$ such that $\mathsf{x}_s = \bar{\mathsf{x}} + t_s \mathsf{v}_s \in M_{\bar{\mathsf{u}}}$. Then we have by continuity also $\mathsf{v}_s^\top D^2_{\mathsf{x}} L(\bar{\mathsf{x}}; \bar{\mathsf{u}})\mathsf{v}_s < 0$ if $s > 0$ is small enough, and therefore

$$f(\mathsf{x}_s) = L(\mathsf{x}_s; \bar{\mathsf{u}}) = L(\bar{\mathsf{x}}; \bar{\mathsf{u}}) + \tfrac{t_s^2}{2}\, \mathsf{v}_s^\top D^2_{\mathsf{x}} L(\bar{\mathsf{x}}; \bar{\mathsf{u}})\mathsf{v}_s + \mathfrak{o}(t_s^2) < f(\bar{\mathsf{x}})$$

if $s > 0$ is small enough, contradicting local optimality of $\bar{\mathsf{x}}$. □

Remark 1.3 *The necessary second-order conditions can also be satisfied if RACQ fails: this example is due to F. Facchinei, A. Fischer (personal communication) and M. Herrich who adapted [7, Ex.2.2], see also the references therein; also cf. [5, Ex.2.2]: Let* $n = 3$, $m = 2$, $q = 0$ *and consider the non-convex problem given by* $f(\mathsf{x}) = x_1^2 - x_2^2 + x_3^2$ *and* $G(\mathsf{x}) = [x_1^2 + x_2^2 - x_3^2, x_1 x_3]^\top$. *Obviously,* $f(\mathsf{x}) \geq 2x_1^2 \geq 0$ *on the feasible set (which is unbounded), so* $\mathsf{x}^* = \mathsf{o}$ *is optimal. Further, the Lagrangian has derivatives*

$$\nabla_{\mathsf{x}} L(\mathsf{x}; \mathsf{u}) = \begin{bmatrix} 2(1+u_1)x_1 + u_2 x_3 \\ 2(u_1 - 1)x_2 \\ 2(1-u_1)x_3 + u_2 x_1 \end{bmatrix}, \; D^2_{\mathsf{x}} L(\mathsf{x}; \mathsf{u}) = 2\begin{bmatrix} 1+u_1 & 0 & \frac{u_2}{2} \\ 0 & u_1 - 1 & 0 \\ \frac{u_2}{2} & 0 & 1-u_1 \end{bmatrix}.$$

We conclude there is a continuum (in fact, two branches) of further optimal solutions $\mathsf{x}_t^\pm = [0, t, \pm t]^\top$ *as* $t \neq 0$ *at which the dual variables* u_t *are unique, since they all equal* $\mathsf{u}^* = [1, 0]^\top$, *while at* x^*, *any* $\mathsf{u} \in \mathbb{R}^2_+$ *satisfy the KKT system* $\nabla_{\mathsf{x}} L(\mathsf{x}^*; \mathsf{u}) = \mathsf{o}$. *It is easy to see that there are no other KKT points for this problem. Next we investigate*

$$D_{\mathsf{x}} G(\mathsf{x}_t^\pm) = \begin{bmatrix} 0 & 2t & \pm 2t \\ \pm t & 0 & 0 \end{bmatrix}$$

to see that the LICQ are satisfied for $t \neq 0$ while they fail at x^* (just put $t = 0$). Now for $t = 0$ (i.e., at x^*) as well for all other t, we have

$$\Gamma(\mathsf{x}_t^{\pm}) = \left\{\mathsf{v} \in \mathbb{R}^3 : tv_2 \leq \pm tv_3 \text{ and } \pm tv_1 \leq 0\right\} ,$$

which means $\Gamma(\mathsf{x}^*) = \mathbb{R}^3$. Further, the reduced tangent cones are $\Gamma_0(\mathsf{x}^*) = \mathbb{R}^3$ still while

$$\Gamma_0(\mathsf{x}_t^{\pm}) = \begin{cases} \left\{\mathsf{v} \in \mathbb{R}^3 : \pm v_1 \leq 0 \text{ and } v_2 = \pm v_3\right\} , & \text{if } t > 0 , \\ \left\{\mathsf{v} \in \mathbb{R}^3 : \pm v_1 \leq 0 \text{ and } v_2 = \pm v_3\right\} , & \text{if } t < 0 . \end{cases}$$

Anyhow, we see that $D^2_{\mathsf{x}} L(\mathsf{x}_t^{\pm}; \mathsf{u}^*)$ is positive-semidefinite for all t and therefore $\Gamma_0(\mathsf{x}_t^{\pm})$-copositive (including $t = 0$, i.e., also at x^*), while for all other choices of u, the Hessian $D^2_{\mathsf{x}} L(\mathsf{x}^*; \mathsf{u})$ is not *$\Gamma_0(\mathsf{x}^*)$-copositive. Nevertheless, even at x^*, RACQ is clearly violated for*

$$M_{\mathsf{u}^*} = \left\{\mathsf{x} \in \mathbb{R}^3 : x_1^2 + x_2^2 = x_3^2, x_1 x_3 \leq 0\right\} ,$$

i.e., for $(G, H) = (g_2, g_1)$. Note that x^* is not an isolated (local) solution to the problem, and that GCQ is satisfied for M_{u^*} at x^* as $[T_{M_{\mathsf{u}^*}}(\mathsf{x}^*)]^{\star} = [T_{M_{\mathsf{u}^*}}(\mathsf{x}^*)]^{\perp} = \{\mathsf{o}\} = [\Gamma_0(\mathsf{x}^*)]^{\star}$. This example will be analyzed further in Section 4 dealing with the SDP relaxation of all-quadratic problems of this type.

1.3 Special case: quadratic optimization over polyhedra (QP)

Theorem 1.1 says

$$\text{strict copositivity} \Rightarrow \text{strict local solution} \Rightarrow \text{local solution} \Rightarrow \text{copositivity} , \tag{3}$$

and for quadratic optimization problems over polyhedra, the leftmost and the rightmost implications in (3) become equivalences. This has been known before, see, e.g. [2, p.5] and the references therein. But in the strict case, one can specify an explicit error bound (implying even strong optimality rather than strict optimality), and this will done below (apparently for the first time in literature).

For ease of particular reference, we provide a separate proof also for the implications which already have been established in Theorem 1.1. Beforehand we note that both f and L are quadratic functions so that the Taylor expansions of order two are exact for both functions, and that their Hessians coincide: $D^2_{\mathsf{x}} f(\mathsf{x}) = D^2_{\mathsf{x}} L(\mathsf{x}; \mathsf{u}) = \mathsf{Q}$, a constant matrix. Further note that no constraint qualifications are needed in this case.

We need the following auxiliary result.

Lemma 1.2 *Let Γ be a polyhedral cone, $\mathsf{Q} = \mathsf{Q}^\top$ an $n \times n$ matrix and $\mathsf{c} \in \mathbb{R}^n$. Suppose $\mathsf{c} \in \Gamma^\star$ and denote by $\Gamma_0 = \Gamma \cap \mathsf{c}^\perp$.*
(a) If Q is strictly Γ_0-copositive, then there is an $\varepsilon > 0$ and a $\rho > 0$ such that

$$\mathsf{c}^\top \mathsf{v} + \frac{1}{2}\mathsf{v}^\top \mathsf{Q}\mathsf{v} \geq \rho \|\mathsf{v}\|^2 \quad \textit{for all } \mathsf{v} \in \Gamma \textit{ with } \|\mathsf{v}\| < \varepsilon\,;$$

(b) If Q is Γ_0-copositive, then there is an $\varepsilon > 0$ such that

$$\mathsf{c}^\top \mathsf{v} + \frac{1}{2}\mathsf{v}^\top \mathsf{Q}\mathsf{v} \geq 0 \quad \textit{for all } \mathsf{v} \in \Gamma \textit{ with } \|\mathsf{v}\| < \varepsilon\,.$$

Proof. (a) Since Γ is polyhedral, it is generated by finitely many extremal rays, i.e., there are $\mathsf{r}_1, \ldots, \mathsf{r}_k$ with $\|\mathsf{r}_i\| = 1$ such that $\Gamma = \mathbb{R}_+\mathrm{conv}\,(\mathsf{r}_1, \ldots, \mathsf{r}_k)$. Further, define $B_0 = \{\mathsf{v} \in \Gamma_0 : \|\mathsf{v}\| = 1\}$; then, by assumption, we have $\delta := \min\limits_{\mathsf{v}\in B} \mathsf{v}^\top \mathsf{Q}\mathsf{v} > 0$. First, consider the case that $\Gamma \subseteq \mathsf{c}^\perp$ so that $\Gamma_0 = \Gamma$ and for $\rho = \frac{\delta}{2}$

$$\mathsf{c}^\top \mathsf{v} + \frac{1}{2}\mathsf{v}^\top \mathsf{Q}\mathsf{v} = \frac{1}{2}\mathsf{v}^\top \mathsf{Q}\mathsf{v} \geq \rho \|\mathsf{v}\|^2$$

even for all $\mathsf{v} \in \Gamma$ regardless of their norm. A bit more care is required if $\Gamma_0 \neq \Gamma$. Assume without loss of generality that $\mathsf{c}^\top \mathsf{r}_i = 0$ for $1 \leq i \leq s$ and $\mathsf{c}^\top \mathsf{r}_i > 0$ for $s < i \leq k$. By rescaling everything, we may and do also assume that $\|\mathsf{c}\| = 1$. Hence $\bar{\mathsf{r}}_i = (\mathsf{c}^\top \mathsf{r}_i)\mathsf{c}$ is the orthoprojection of r_i onto $\mathbb{R}\mathsf{c}$. Next, for any $\mathsf{v} \in \Gamma$ there are $\mu_i \geq 0$ such that $\mathsf{v} = \sum_{i=1}^k \mu_i \mathsf{r}_i$ and we define

$$\begin{aligned}
\mathsf{w} &= \sum_{i=s+1}^k \mu_i \bar{\mathsf{r}}_i = \left(\sum_{i=s+1}^k \mu_i \mathsf{c}^\top \mathsf{r}_i\right)\mathsf{c}\,,\\
\mathsf{y} &= \sum_{i=s+1}^k \mu_i (\mathsf{r}_i - \bar{\mathsf{r}}_i)\,,\\
\mathsf{z} &= \sum_{i=1}^k \mu_i \mathsf{r}_i \in \Gamma_0\,.
\end{aligned}$$

This way we obtained a decomposition $\mathsf{v} = \mathsf{w} + \mathsf{y} + \mathsf{z}$ with $\mathsf{w} = \|\mathsf{w}\|\mathsf{c}$ orthogonal to $\mathsf{y} + \mathsf{z}$. Indeed, we have $\mathsf{c}^\top \mathsf{y} = \sum\limits_{i>s} \mu_i(\mathsf{c}^\top \mathsf{r}_i - \|\mathsf{c}\|^2 \mathsf{c}^\top \mathsf{r}_i) = 0$, so $\mathsf{y} \perp \mathsf{c}$ and furthermore

$$\left.\begin{aligned}
\|\mathsf{y}\| &\leq \sum_{i>s} \mu_i \|\mathsf{r}_i - \bar{\mathsf{r}}_i\| = \sum_{i>s} \mu_i \sqrt{1 - (\mathsf{c}^\top \mathsf{r}_i)^2}\\
&\leq \sum_{i>s} \mu_i \eta (\mathsf{c}^\top \mathsf{r}_i) = \eta \|\mathsf{w}\|\,,
\end{aligned}\right\} \tag{4}$$

where $\eta = \max\limits_{s<i\leq k} \frac{\sqrt{1-(\mathsf{c}^\top \mathsf{r}_i)^2}}{\mathsf{c}^\top \mathsf{r}_i}$. The first equality above follows by

$$\|\mathsf{r}_i - \bar{\mathsf{r}}_i\|^2 = \|\mathsf{r}_i\|^2 - \|\bar{\mathsf{r}}_i\|^2 = 1 - (\mathsf{c}^\top \mathsf{r}_i)^2 \|\mathsf{c}\|^2 = 1 - (\mathsf{c}^\top \mathsf{r}_i)^2\,.$$

We note for later use that this implies

$$\left.\begin{array}{lll} \|\mathsf{v}\| & \leq & (1+\eta)\|\mathsf{w}\|+\|\mathsf{z}\| \quad \text{and thus} \\ \|\mathsf{v}\|^2 & \leq & (1+\eta)(2+\eta)\|\mathsf{w}\|^2+(2+\eta)\|\mathsf{z}\|^2 \,. \end{array}\right\} \tag{5}$$

Next choose a number $\beta > 0$ such that $|\mathsf{p}^\top \mathsf{Q}\mathsf{q}| \leq 2\beta\|\mathsf{p}\|\,\|\mathsf{q}\|$ for all $\{\mathsf{p},\mathsf{q}\} \subset \mathbb{R}^n$. It follows from $\mathsf{v} = \mathsf{w}+\mathsf{y}+\mathsf{z}$, from $\mathsf{c}^\top\mathsf{w} = \|\mathsf{w}\|$ and from (4) that

$$\left.\begin{array}{ll} & \mathsf{c}^\top\mathsf{v}+\frac{1}{2}\mathsf{v}^\top\mathsf{Q}\mathsf{v} \\ = & \mathsf{c}^\top\mathsf{w}+\frac{1}{2}\mathsf{w}^\top\mathsf{Q}\mathsf{w}+\mathsf{w}^\top\mathsf{Q}\mathsf{y}+\mathsf{w}^\top\mathsf{Q}\mathsf{z}+\mathsf{y}^\top\mathsf{Q}\mathsf{z}+\frac{1}{2}\mathsf{y}^\top\mathsf{Q}\mathsf{y}+\frac{1}{2}\mathsf{z}^\top\mathsf{Q}\mathsf{z} \\ \geq & \|\mathsf{w}\|-\beta\|\mathsf{w}\|^2-2\beta\|\mathsf{w}\|\,(\|\mathsf{y}\|+\|\mathsf{z}\|)-2\beta\|\mathsf{y}\|\,\|\mathsf{z}\|-\beta\|\mathsf{y}\|^2+\delta\|\mathsf{z}\|^2 \\ \geq & \|\mathsf{w}\|\,\big(1-\beta\|\mathsf{w}\|-2\beta\eta\|\mathsf{w}\|-2\beta\|\mathsf{z}\|-2\beta\eta\|\mathsf{z}\|-\beta\eta^2\|\mathsf{w}\|\big)+\delta\|\mathsf{z}\|^2 \\ = & \|\mathsf{w}\|\,\big(1-\beta(1+\eta)^2\|\mathsf{w}\|-2\beta(1+\eta)\|\mathsf{z}\|\big)+\delta\|\mathsf{z}\|^2 \,. \end{array}\right\} \tag{6}$$

Now $\mathsf{w} = \|\mathsf{w}\|\mathsf{c}$ is orthogonal to $\mathsf{y}+\mathsf{z}$, so $\|\mathsf{v}\|^2 = \|\mathsf{w}\|^2+\|\mathsf{y}+\mathsf{z}\|^2$. Hence the inequality relation $\|\mathsf{v}\| \leq \varepsilon$ entails both $\|\mathsf{w}\| \leq \varepsilon$ and also $\|\mathsf{y}+\mathsf{z}\| \leq \varepsilon$. Therefore also

$$\|\mathsf{z}\| \leq \|\mathsf{y}+\mathsf{z}\|+\|\mathsf{y}\| \leq \varepsilon+\eta\|\mathsf{w}\| \leq \varepsilon(1+\eta)\,. \tag{7}$$

Now the factor of $\|\mathsf{w}\|$ in the last line of (6) exceeds $(1+\eta)\delta\|\mathsf{w}\|$ if

$$\big[(1+\eta)\delta+(1+\eta)^2\beta\big]\,\|\mathsf{w}\|+2(1+\eta)\beta\|\mathsf{z}\| \leq 1\,,$$

and this can in turn be achieved if $\varepsilon > 0$ is selected so small that

$$(1+\eta)\delta+3(1+\eta)^2\beta \leq \frac{1}{\varepsilon}\,,$$

using the relations $\|\mathsf{w}\| \leq \varepsilon$ and $\|\mathsf{z}\| \leq \varepsilon(1+\eta)$ established in (7). Then we arrive via (6) and (5) at

$$\mathsf{c}^\top\mathsf{v}+\tfrac{1}{2}\,\mathsf{v}^\top\mathsf{Q}\mathsf{v} \geq (1+\eta)\delta\|\mathsf{w}\|^2+\delta\|\mathsf{z}\|^2 \geq \tfrac{\delta}{2+\eta}\,\|\mathsf{v}\|^2\,.$$

Claim (a) is proved by putting $\rho = \frac{\delta}{2+\eta}$ and, e.g., $\varepsilon = [4(1+\eta)^2\max\{\delta,\beta\}]^{-1}$. Assertion (b) can be found in [1, Lemma 1]. The proof of (a) above is a refinement of the arguments there. □

Theorem 1.2 *Let* $f(\mathsf{x}) = \frac{1}{2}\mathsf{x}^\top\mathsf{Q}\mathsf{x}+\mathsf{q}^\top\mathsf{x}$ *be quadratic and* M *be a polyhedron, and suppose that* $\bar{\mathsf{x}}$ *be a KKT point. Then*

(a) Q *is strictly* $\Gamma_0(\bar{\mathsf{x}})$*-copositive if and only if* $\bar{\mathsf{x}}$ *is a strict local solution;*

(b) Q *is* $\Gamma_0(\bar{\mathsf{x}})$*-copositive if and only if* $\bar{\mathsf{x}}$ *is a local solution.*

Proof. We first show that (strict) copositivity implies (strict) optimality. Let $\mathsf{c} = \nabla f(\bar{\mathsf{x}}) = \mathsf{Q}\bar{\mathsf{x}} + \mathsf{q} \in \Gamma^\star(\bar{\mathsf{x}})$, and apply Lemma 1.2 to $\mathsf{Q} = D^2_{\mathsf{x}} f(\bar{\mathsf{x}})$. For any $\mathsf{x} \in M$ with $\|\mathsf{x} - \bar{\mathsf{x}}\| < \varepsilon$, let $\mathsf{v} = \mathsf{x} - \bar{\mathsf{x}}$. Then by convexity of the polyhedron M we get $\mathsf{v} \in \Gamma(\bar{\mathsf{x}})$ and $\|\mathsf{v}\| < \varepsilon$. We conclude

$$f(\mathsf{x}) = f(\bar{\mathsf{x}}) + \mathsf{c}^\top \mathsf{v} \; + \tfrac{1}{2}\, \mathsf{v}^\top \mathsf{Q} \mathsf{v} \geq f(\bar{\mathsf{x}}) + r(\mathsf{v})\,,$$

where $r(\mathsf{v}) = \rho\|\mathsf{v}\|^2$ in case of strict copositivity while $r(\mathsf{v}) = 0$ for the merely copositive case.

Next we show the converse: (strict) optimality implies (strict) copositivity. So suppose that $\mathsf{v} \in \Gamma_0(\bar{\mathsf{x}}) \setminus \{\mathsf{o}\}$, in particular that $\mathsf{v}^\top \nabla f(\bar{\mathsf{x}}) = 0$. We infer that for small enough $t > 0$, we have $\mathsf{x} = \bar{\mathsf{x}} + t\mathsf{v} \in M$ since M is a polyhedron, and $0 < \|\mathsf{x} - \bar{\mathsf{x}}\| < \varepsilon$, so that

$$f(\bar{\mathsf{x}}) < f(\mathsf{x}) = f(\bar{\mathsf{x}}) + t\, \mathsf{v}^\top \nabla f(\bar{\mathsf{x}}) + \tfrac{t^2}{2}\, \mathsf{v}^\top \mathsf{Q} \mathsf{v} = f(\bar{\mathsf{x}}) + \tfrac{t^2}{2}\, \mathsf{v}^\top \mathsf{Q} \mathsf{v}$$

in case of a strict local solution, or weak inequality in case of a local solution, which implies (strict) $\Gamma_0(\bar{\mathsf{x}})$-copositivity of Q, as claimed. □

2 Second-order conditions for global optimality

2.1 The general case: a sufficient global optimality condition

We return to the general non-linear case and proceed to a sufficient global optimality condition. This condition is weaker than convexity of the Lagrangian function $L(.;\bar{\mathsf{u}})$ which in turn would be implied by the convexity of the problem (1) (which means that f and all g_i are convex and h_i are affine-linear). Nevertheless it requires checking copositivity of the Hessian matrices $D^2_{\mathsf{x}} L(\mathsf{x};\bar{\mathsf{u}})$ for all $\mathsf{x} \in M$, which may be tedious unless we know that these Hessians do not depend on x. So the quadratic problems over polyhedra studied in Theorem 1.2 provide a good motivation to consider this case, but also, more generally, the case where all f, G, H are composed of quadratic functions.

Theorem 2.1 *Suppose that* M *is convex. If* $\bar{\mathsf{x}}$ *is a KKT point for problem (1) with multipliers* $\bar{\mathsf{u}}$ *and if*

$$D^2_{\mathsf{x}} L(\mathsf{x};\bar{\mathsf{u}}) \quad \text{is } \Gamma(\bar{\mathsf{x}})\text{-copositive for all } \mathsf{x} \in M\,,$$

then $\bar{\mathsf{x}}$ *is a global solution to (1).*

Proof. For any $\mathsf{x} \in M$, define the trajectory $\mathsf{z}(t) = (1-t)\bar{\mathsf{x}} + t\mathsf{x} \in M$ (so that $\mathsf{v} = \mathsf{x} - \bar{\mathsf{x}} = \frac{1}{t}[\mathsf{z}(t) - \mathsf{z}(0)] \in \Gamma(\bar{\mathsf{x}})$ as before), as well as the function $\varphi(t) = L(\mathsf{z}(t); \mathsf{u})$ for $0 \le t \le 1$. Now φ is twice continuously differentiable and by the Mean Value Theorem there is some t with $0 < t < 1$ such that

$$
\begin{aligned}
f(\mathsf{x}) \ge L(\mathsf{x}; \bar{\mathsf{u}}) = \varphi(1) &= \varphi(0) + \dot{\varphi}(0) + \tfrac{1}{2}\ddot{\varphi}(t) \\
&= L(\bar{\mathsf{x}}; \bar{\mathsf{u}}) + \mathsf{v}^\top L(\bar{\mathsf{x}}; \bar{\mathsf{u}}) + \tfrac{1}{2}\, \mathsf{v}^\top D_{\mathsf{x}}^2 L(\mathsf{z}(t); \bar{\mathsf{u}})\mathsf{v} \\
&\ge L(\bar{\mathsf{x}}; \bar{\mathsf{u}}) = f(\bar{\mathsf{x}}) \,,
\end{aligned}
$$

and the assertion is shown. □

2.2 Second-order global optimality criterion for QP case

Again, for the QP case, the situation is much simpler, and necessary and sufficient conditions for global optimality coincide. Consider again

$$
\min \left\{ f(\mathsf{x}) = \tfrac{1}{2}\, \mathsf{x}^\top \mathsf{Q}\mathsf{x} + \mathsf{q}^\top \mathsf{x} : \mathsf{x} \in M \right\} \,, \tag{8}
$$

with Q a symmetric $n \times n$ matrix, $M = \{\mathsf{x} \in \mathbb{R}^n : \mathsf{A}\mathsf{x} \le \mathsf{b}\}$, A an $m \times n$ matrix with rows $\mathsf{a}_i^\top$, and $\mathsf{b} \in \mathbb{R}^m$. Due to linearity of the constraints, we have

$$
\Gamma(\bar{\mathsf{x}}) = \mathbb{R}_+(M - \bar{\mathsf{x}}) = \left\{ \mathsf{v} \in \mathbb{R}^n : \mathsf{a}_i^\top \mathsf{v} \le 0 \text{ for all } i \in I(\bar{\mathsf{x}}) \right\} . \tag{9}
$$

Denote by $\mathsf{s} = \mathsf{b} - \mathsf{A}\bar{\mathsf{x}}$ the vector of slack variables, and by $J(\bar{\mathsf{x}}) = \{0, \dots, m\} \setminus I(\bar{\mathsf{x}})$. To have a consistent notation, we can also view as $J(\bar{\mathsf{x}})$ as the set of inactive constraints if we add an auxiliary inactive constraint of the form $0 < 1$ by enriching (A, b) with a 0-th row to

$$
\bar{\mathsf{A}} = \begin{bmatrix} \mathsf{a}_0^\top \\ \\ \mathsf{A} \\ \\ \end{bmatrix} = \begin{bmatrix} \mathsf{o}^\top \\ \mathsf{a}_1^\top \\ \vdots \\ \mathsf{a}_m^\top \end{bmatrix} , \quad \bar{\mathsf{b}} = \begin{bmatrix} b_0 \\ \\ \mathsf{b} \\ \\ \end{bmatrix} = \begin{bmatrix} 1 \\ b_1 \\ \vdots \\ b_m \end{bmatrix} ,
$$

and put $\bar{\mathsf{s}} = \bar{\mathsf{b}} - \bar{\mathsf{A}}\bar{\mathsf{x}} \ge \mathsf{o}$. Then $J(\bar{\mathsf{x}}) = \{i \in \{0, \dots, m\} : \bar{s}_i > 0\}$. The 0-th slack and the corresponding constraint will be needed for dealing with unbounded feasible directions. However, if v is a bounded feasible direction, then there is an $i \in J(\bar{\mathsf{x}}) \setminus \{0\}$ such that $\mathsf{a}_i^\top(\bar{\mathsf{x}} + t\mathsf{v}) > b_i$ for some $t > 0$, and the *maximal feasible stepsize* in direction of v

$$
\bar{t}_{\mathsf{v}} = \min \left\{ \frac{\bar{s}_i}{\bar{\mathsf{a}}_i^\top \mathsf{v}} : i \in J(\bar{\mathsf{x}}) \,,\ \bar{\mathsf{a}}_i^\top \mathsf{v} > 0 \right\}
$$

is finite.

Note that feasibility of a direction $\mathsf{v} \in \mathbb{R}^n$ is fully characterized by the property $\mathsf{a}_i^\top \mathsf{v} \leq 0$ for all i with $s_i = 0$, i.e., for all i in the complement of $J(\bar{\mathsf{x}})$. If in addition, $\mathsf{a}_i^\top \mathsf{v} \leq 0$ for all $i \in J(\bar{\mathsf{x}})$, i.e., $\mathsf{A}\mathsf{v} \leq \mathsf{o}$ but $\mathsf{v} \neq \mathsf{o}$, then we have an unbounded feasible direction with $\bar{t}_{\mathsf{v}} = +\infty$ by the usual default rules, consistent with the property that $\bar{\mathsf{x}} + t\mathsf{v} \in M$ for all $t > 0$ in this case. In the opposite case where $\bar{t}_{\mathsf{v}} = \frac{s_i}{\mathsf{a}_i^\top \mathsf{v}} < +\infty$, we have $i \neq 0$, and the i-th constraint is the first inactive constraint which becomes active when travelling from $\bar{\mathsf{x}}$ along the ray given by v: then $\bar{\mathsf{x}} + \bar{t}_{\mathsf{v}}\mathsf{v} \in M$, but $\bar{\mathsf{x}} + t\mathsf{v} \notin M$ for all $t > \bar{t}_{\mathsf{v}}$.

By consequence, the feasible polyhedron M is decomposed into a union of polytopes $M_i(\bar{\mathsf{x}}) = \{\mathsf{x} = \bar{\mathsf{x}} + t\mathsf{v} \in M : 0 \leq t \leq \bar{t}_{\mathsf{v}}\}$ for $i \in J(\bar{\mathsf{x}}) \setminus \{0\}$, and $M_0(\bar{\mathsf{x}}) = \{\mathsf{x} \in \mathbb{R}^n : \mathsf{A}\mathsf{x} \leq \mathsf{A}\bar{\mathsf{x}}\}$, the (possibly trivial, but otherwise) unbounded polyhedral part of M.

To be more precise, we need the $(m+1) \times n$-matrices $\mathsf{D}_i = \bar{\mathsf{s}}\,\mathsf{a}_i^\top - \bar{s}_i \bar{\mathsf{A}}$ to define the polyhedral cones

$$\Gamma_i = \{\mathsf{v} \in \mathbb{R}^n : \mathsf{D}_i \mathsf{v} \geq \mathsf{o}\} \ , \quad i \in J(\bar{\mathsf{x}}) \, . \tag{10}$$

Then $\bigcup_{i \in J(\bar{\mathsf{x}})} \Gamma_i = \Gamma(\bar{\mathsf{x}})$ from (9), and $M_i(\bar{\mathsf{x}}) = M \cap (\Gamma_i + \bar{\mathsf{x}})$ contains all points $\mathsf{x} \in M$ where $i \in J(\bar{\mathsf{x}})$ denotes the first inactive constraint which becomes active when travelling along direction $\mathsf{x} - \bar{\mathsf{x}}$ starting from $\bar{\mathsf{x}}$ (as mentioned above, the case $i = 0$ captures unbounded feasible directions).

After these preparations dealing with the feasible set only, we turn to the objective function. With the gradient $\nabla f(\bar{\mathsf{x}}) = \mathsf{Q}\bar{\mathsf{x}} + \mathsf{q}$, we construct rank-two updates of Q:

$$\mathsf{Q}_i = \mathsf{a}_i \, \nabla f(\bar{\mathsf{x}})^\top + \nabla f(\bar{\mathsf{x}}) \, \mathsf{a}_i^\top + \bar{s}_i \mathsf{Q} \, , \quad i \in J(\bar{\mathsf{x}}) \, . \tag{11}$$

Theorem 2.2 *For the QP (8), we have that $\bar{\mathsf{x}}$ is a global solution to (8) if and only if $\bar{\mathsf{x}}$ is a KKT point and*

$$\mathsf{Q}_i \text{ are } \Gamma_i\text{-copositive} \quad \text{for all } i \in J(\bar{\mathsf{x}}) \, .$$

Else, if $\mathsf{v}^\top \mathsf{Q}_i \mathsf{v} < 0$ and $\mathsf{D}_i \mathsf{v} \geq \mathsf{o}$ for some $i \in J(\bar{\mathsf{x}}) \setminus \{0\}$, then $\mathsf{a}_i^\top \mathsf{v} > 0$ and

$$\widetilde{\mathsf{x}} = \bar{\mathsf{x}} + \bar{t}_{\mathsf{v}} \mathsf{v} \quad \text{is an improving feasible point,}$$

whereas $\mathsf{v}^\top \mathsf{Q}_0 \mathsf{v} < 0$ for some v with $\mathsf{D}_0 \mathsf{v} \geq \mathsf{o}$ if and only if (8) is unbounded.

Proof. As already noted before, any KKT point satisfies the weak first-order ascent condition $\mathsf{v}^\top \nabla f(\bar{\mathsf{x}}) \geq 0$ for all $\mathsf{v} \in \Gamma(\bar{\mathsf{x}})$. Further, strict first-order ascent directions may be negative curvature directions as

$$f(\bar{\mathsf{x}} + t\mathsf{v}) - f(\bar{\mathsf{x}}) = t \left[\mathsf{v}^\top \nabla f(\bar{\mathsf{x}}) + \tfrac{t}{2} \, \mathsf{v}^\top \mathsf{Q} \mathsf{v} \right] > 0 \, , \tag{12}$$

if $t > 0$ is small enough and $\mathsf{v}^\top \nabla f(\bar{\mathsf{x}}) > 0$, even if $\mathsf{v}^\top \mathsf{Q} \mathsf{v} < 0$. For these negative curvature directions ($\mathsf{v}^\top \mathsf{Q} \mathsf{v} < 0$), the *extremal increment*

$$\theta_{\bar{\mathsf{x}}}(\mathsf{v}) = f(\bar{\mathsf{x}} + \bar{t}_{\mathsf{v}} \mathsf{v}) - f(\bar{\mathsf{x}})$$

satisfies, due to (12),

$$f(\bar{\mathsf{x}} + t\mathsf{v}) - f(\bar{\mathsf{x}}) \geq 0 \quad \text{for all } \mathsf{x} = \bar{\mathsf{x}} + t\mathsf{v} \in M \text{ , i.e., } t \in [0, \bar{t}_{\mathsf{v}}] \, ,$$

if and only if $\theta_{\bar{\mathsf{x}}}(\mathsf{v}) \geq 0$. For $\mathsf{v} \in \Gamma_i$, the condition $\theta_{\bar{\mathsf{x}}}(\mathsf{v}) \geq 0$ can be expressed as $\mathsf{v}^\top \mathsf{Q}_i \mathsf{v} \geq 0$. Hence the result. □

Note that the result above also applies to QPs for which the Frank/Wolfe-Theorem is non-trivial, i.e., where the objective function f is bounded from below over an unbounded polyhedron M.

Comparing Theorems 1.2 and 2.2, we see that the effort of checking local versus global optimality is not that different: at most m copositivity checks instead of merely one. Also note that any vector v violating the copositivity conditions in Theorem 2.2 yields with basically no effort an improving feasible point $\widetilde{\mathsf{x}}$, hence allows for escaping from inefficient local solutions $\bar{\mathsf{x}}$ towards which a local optimization procedure may have driven us before.

Remark 2.1 *Of course, Theorems 2.1 and 2.2 immediately yield via global optimality that $\Gamma(\bar{\mathsf{x}})$-copositivity of $D^2_{\mathsf{x}} L(\mathsf{x}; u) = \mathsf{Q}$ implies that all Q_i are Γ_i-copositive. However it may be instructive to see how the copositivity conditions are related directly. To this end, observe that the* 0*-th row of D_i in (10) equals $\mathsf{a}_i^\top$, so that $\mathsf{a}_i^\top \mathsf{v} \geq 0$ holds for all $\mathsf{v} \in \Gamma_i$. Further, any $\mathsf{v} \in \Gamma_i \subseteq \Gamma(\bar{\mathsf{x}})$ satisfies $\mathsf{v}^\top \nabla f(\bar{\mathsf{x}}) \geq 0$. Therefore (11) renders*

$$\mathsf{v}^\top \mathsf{Q}_i \mathsf{v} \geq \bar{s}_i \, \mathsf{v}^\top \mathsf{Q} \mathsf{v} \geq 0 \quad \textit{for all } \mathsf{v} \in \Gamma_i \, .$$

On the other hand, also the local optimality condition of Theorem 1.2 can be retrieved readily in a direct way, noticing that $\mathsf{v}^\top \mathsf{Q}_i \mathsf{v} = \bar{s}_i \, \mathsf{v}^\top \mathsf{Q} \mathsf{v}$ for all $\mathsf{v} \perp \nabla f(\bar{\mathsf{x}})$ and the fact that $\Gamma(\bar{\mathsf{x}}) = \bigcup_{i \in J(\bar{\mathsf{x}})} \Gamma_i$.

3 Application: the classical trust region problem

3.1 Definition and basic properties

Now we specialize our findings to the well studied classical trust region problem where $f(\mathsf{x}) = \frac{1}{2} \mathsf{x}^\top \mathsf{Q} \mathsf{x} + \mathsf{q}^\top \mathsf{x}$ is quadratic and the feasible set M is the (convex) Euclidean ball centered at the origin with radius one:

$$\min\left\{ f(\mathsf{x}) = \tfrac{1}{2}\,\mathsf{x}^\top \mathsf{Q}\mathsf{x} + \mathsf{q}^\top \mathsf{x} : \mathsf{x} \in \mathbb{R}^n\,,\ \|\mathsf{x}\| \le 1 \right\}\,. \tag{13}$$

All results in this section are well known since quite a while; however, for illustration we derive them from our copositivity principles established above. Thus, we have $m = 1$ inequality constraint $g_1(\mathsf{x}) = r(\mathsf{x}) = \frac{1}{2}(\|\mathsf{x}\|^2 - 1)$ and no equality constraints. If $r(\mathsf{x}) = 0$, then the gradient $\nabla r(\mathsf{x}) = \mathsf{x}$ is linearly independent, so LICQ (and therefore RACQ) holds in any case. We conclude that all local solutions must be KKT points x satisfying $(\mathsf{Q} + u\mathsf{I}_n)\mathsf{x} = -\mathsf{q}$ for some $u \ge 0$, with $u = 0$ if $\|\mathsf{x}\| < 1$ or else $u = \bar{u} := -\mathsf{x}^\top \mathsf{Q}\mathsf{x} - \mathsf{q}^\top \mathsf{x}$, so u is uniquely determined by x. The Lagrangian function reads

$$L(\mathsf{x}; u) = \tfrac{1}{2}\,\mathsf{x}^\top (\mathsf{Q} + u\mathsf{I}_n)\mathsf{x} + \mathsf{q}^\top \mathsf{x} - \tfrac{u}{2}$$

and has a Hessian $\mathsf{H}_u = \mathsf{Q} + u\mathsf{I}_n$ which does not depend on x.

3.2 Application of second-order optimality conditions

We now illustrate a simple application of preceding general principles to the trust-region problem. First we show that zero multipliers imply global optimality for this problem.

Corollary 3.1 *Suppose* x *is a local solution to (13) (and hence a KKT point) with a multiplier* $u = 0$. *Then* x *is globally optimal. In fact, then* f *is a convex function because* Q *is indeed positive-semidefinite.*

Proof. If $u = 0$, then $M_u = M$ and $\nabla f(\mathsf{x}) = \mathsf{o}$ as the KKT conditions are satisfied. Therefore $\Gamma_0(\mathsf{x}) = \Gamma(\mathsf{x})$, and all constraint qualifications hold as detailed above (these are only necessary if strict complementarity is violated, i.e., if $r(\mathsf{x}) = u = 0$), Thus Theorem 1.1(b) implies $\Gamma(\bar{\mathsf{x}})$-copositivity of $\mathsf{H}_0(\mathsf{x}) = \mathsf{Q}$. Now for $\|\mathsf{x}\| < 1$ we have $\Gamma(\mathsf{x}) = \mathbb{R}^n$ and the result follows. If however $r(\mathsf{x}) = 0$, then $\Gamma(\bar{\mathsf{x}}) = \left\{\mathsf{v} \in \mathbb{R}^n : \mathsf{v}^\top \bar{\mathsf{x}} \le 0\right\}$ is a halfspace, and again Q must be positive-semidefinite. □

Hence, any *local non-global (LNG)* solution $\bar{\mathsf{x}}$ must lie on the boundary of M and has a strictly positive multiplier $\bar{u} > 0$. Again,

$$\Gamma(\bar{\mathsf{x}}) = \left\{\mathsf{v} \in \mathbb{R}^n : \mathsf{v}^\top \bar{\mathsf{x}} \le 0\right\}$$

is a half-space and since $\nabla f(\bar{\mathsf{x}}) = -\bar{\mathsf{u}}\bar{\mathsf{x}}$, its boundary hyperplane is $\Gamma_0(\bar{\mathsf{x}}) = \bar{\mathsf{x}}^\perp$. So only boundary KKT points satisfying strict complementarity can be LNGs. We collect further observations on LNGs in the following

Corollary 3.2 *Suppose* $\bar{\mathsf{x}}$ *is a KKT point of (13) with* $\|\bar{\mathsf{x}}\| = 1$ *and unique multiplier* $\bar{u} = -\bar{\mathsf{x}}^\top \mathsf{Q}\bar{\mathsf{x}} - \mathsf{q}^\top \bar{\mathsf{x}} > 0$ *and denote by* $\mathsf{H}_{\bar{u}} = \mathsf{Q} + \bar{u}\mathsf{I}_n$ *the Hessian of the Lagrangian. Denote by* $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ *the ordered eigenvalues of* Q *(counting multiplicities), and by* v_i *the corresponding orthonormal eigenvectors. Then*

(a) If $\bar{\mathsf{x}}$ *is a local solution to (13) then* $\mathsf{v}^\top \mathsf{H}_{\bar{u}}\mathsf{v} \geq 0$ *if* $\mathsf{v}^\top \bar{\mathsf{x}} = 0$*. So* $\mathsf{H}_{\bar{u}}$ *can have at most one negative eigenvalue (and then with multiplicity one).*

(b) If $\bar{\mathsf{x}}$ *is a LNG solution to (13), then* $-\lambda_2 \leq \bar{u} < -\lambda_1$ *and* $\mathsf{v}_1^\top \bar{\mathsf{x}} \neq 0$*.*

(c) Further, if $\bar{\mathsf{x}}$ *is a LNG solution to (13), then* $\mathsf{v}_1^\top \mathsf{q} \neq 0$ *and* $\mathsf{v}_2^\top \bar{\mathsf{x}} \neq 0$*.*

(d) Further, if $\bar{\mathsf{x}}$ *is a LNG solution to (13), then even* $-\lambda_2 < \bar{u} < -\lambda_1$*. In particular,* $\mathsf{H}_{\bar{u}}$ *is nonsingular.*

Proof. (a) We have $M_{\bar{u}} = \partial M = \{\mathsf{x} \in \mathbb{R}^n : \|\mathsf{x}\| = 1\}$ and any $\mathsf{v} \in \bar{\mathsf{x}}^\perp = \Gamma_0(\mathsf{x})$ gives rise to a starting tangent of a trajectory $\mathsf{y}(t) = \frac{1}{\|\bar{\mathsf{x}}+t\mathsf{v}\|}(\bar{\mathsf{x}} + t\mathsf{v}) \in M_{\bar{u}}$. Hence Theorem 1.1(b) applies and yields $\mathsf{v}^\top \mathsf{H}_{\bar{u}}\mathsf{v} \geq 0$ if $\mathsf{v}^\top \bar{\mathsf{x}} = 0$. Suppose there are two linear independent $\mathsf{v}_1, \mathsf{v}_2$ such that all non-trivial linear combinations $\mathsf{v} = \alpha\mathsf{v}_1 + \beta\mathsf{v}_2$ give a negative quadratic form $\mathsf{v}^\top \mathsf{H}_{\bar{u}}\mathsf{v} < 0$. Then, e.g., $\mathsf{v}_1 \perp \bar{\mathsf{x}}$ is absurd, so $\mathsf{v}_1^\top \bar{\mathsf{x}} \neq 0$. Choose $\beta = 1$ and $\alpha = -\frac{\mathsf{v}_2^\top \bar{\mathsf{x}}}{\mathsf{v}_1^\top \bar{\mathsf{x}}}$ to obtain the contradiction $\mathsf{v}^\top \bar{\mathsf{x}} = 0$.
(b) If $\bar{u} \geq -\lambda_1$, then $\mathsf{H}_{\bar{u}} = \mathsf{Q} + \bar{u}\mathsf{I}_n$ would be positive-semidefinite, and Theorem 2.1 would give global optimality of $\bar{\mathsf{x}}$. Hence $\bar{u} < -\lambda_1$. On the other hand, by (a) $\mathsf{H}_{\bar{u}}$ can have at most one negative eigenvalue, so $\bar{u}+\lambda_2 \geq 0$ must hold. Hence $-\lambda_2 \leq \bar{u} < -\lambda_1$. In particular, $\lambda_1 < \lambda_2$. Since $\mathsf{v}_1^\top \mathsf{H}_{\bar{u}}\mathsf{v}_1 = \mathsf{v}_1^\top(\lambda_1 - \lambda_2)\mathsf{v}_1 = (\lambda_1 - \lambda_2) < 0$, assertion (a) implies $\mathsf{v}_1^\top \bar{\mathsf{x}} \neq 0$.
(c) Now we basically follow Martínez' impressive argumentation [9]. First we show that v_1 cannot be orthogonal to q; indeed, suppose the contrary. Now, if $\bar{u} > -\lambda_2$, then $\mathsf{H}_{\bar{u}}$ is nonsingular, and expanding q in terms of the orthonormal basis $\{\mathsf{v}_1, \ldots, \mathsf{v}_n\}$, we see from $\mathsf{H}_{\bar{u}}\mathsf{x} = \mathsf{q} = \sum_{i=1}^{n} \gamma_i \mathsf{v}_i$ where $\gamma_i = \mathsf{v}_i^\top \mathsf{q}$ (hence the assumption means $\gamma_1 = 0$), that

$$\bar{\mathsf{x}} = -\sum_{i=1}^{n} \frac{\gamma_i}{\lambda_i + \bar{u}} \mathsf{v}_i = -\sum_{i=2}^{n} \frac{\gamma_i}{\lambda_i + \bar{u}} \mathsf{v}_i \perp \mathsf{v}_1 ,$$

contradicting assertion (b). Hence $\bar{u} = -\lambda_2$. To be more precise, assume that $\lambda_2 = \ldots = \lambda_k < \lambda_{k+1}$ for some $k \in \{2, \ldots n\}$ (with a trivial modification if $\lambda_2 = \lambda_n$), so that $\{\mathsf{v}_2, \ldots, \mathsf{v}_k\}$ span the null space of $\mathsf{H}_{\bar{u}}$. As $\mathsf{H}_{\bar{u}}\mathsf{x} = -\mathsf{q}$ must hold, it follows

$$\bar{\mathsf{x}} = -\mathsf{H}_{\bar{u}}^{+}\mathsf{q} + \sum_{j=2}^{k} \alpha_j \mathsf{v}_j ,$$

for some suitable coefficients $\alpha_j \in \mathbb{R}$, where, with $\mathsf{V} = [\mathsf{v}_1, \ldots, \mathsf{v}_n]$,

$$\mathsf{H}_{\bar{u}}^{+} = \mathsf{V}\mathrm{diag}\,[\frac{1}{\lambda_1 - \lambda_2}, 0, \ldots, 0, \frac{1}{\lambda_{k+1} - \lambda_2}, \ldots, \frac{1}{\lambda_n - \lambda_2}]\mathsf{V}^\top$$

is the Moore-Penrose generalized inverse of $\mathsf{H}_{\bar{u}}$ (cf. Lemma 4.1 below for details). Again using the assumption $\gamma_1 = \mathsf{v}_1^\top \mathsf{q} = 0$, we arrive at $\mathsf{H}_{\bar{u}}^{+}\mathsf{q} = \sum_{j>k} \frac{\gamma_j}{\lambda_j - \lambda_2} \mathsf{v}_j$, yielding again a contradiction to (b), namely

$$\bar{\mathsf{x}} = \sum_{j=2}^{k} \alpha_j \mathsf{v}_j - \sum_{j>k} \frac{\gamma_j}{\lambda_j - \lambda_2} \mathsf{v}_j \perp \mathsf{v}_1 \,.$$

Hence the assumption is absurd, and we conclude $\mathsf{v}_1^\top \mathsf{q} \neq 0$. Next we prove $\mathsf{v}_2^\top \bar{\mathsf{x}} \neq 0$. Suppose the contrary. We pass again to coordinates w.r.t. the orthonormal basis $\{\mathsf{v}_1, \ldots, \mathsf{v}_n\}$ and consider $\mathsf{w} = \mathsf{V}^\top \mathsf{x}$ instead of x. Of course $\|\mathsf{x}\| = \|\mathsf{w}\|$ and

$$f(\mathsf{x}) = \tfrac{1}{2}\, \mathsf{x}^\top \mathsf{Q} \mathsf{x} + \mathsf{q}^\top \mathsf{x} = \textstyle\sum_{i=1}^{n} [\frac{\lambda_i}{2} w_i^2 + \gamma_i w_i] \,,$$

so that $\overline{\mathsf{w}} = \mathsf{V}^\top \bar{\mathsf{x}}$ is a local solution to the problem

$$\min \left\{ \sum_{i=1}^{n} [\frac{\lambda_i}{2} w_i^2 + \gamma_i w_i] : \|\mathsf{w}\| = 1 \right\} .$$

The assumption $\mathsf{v}_2^\top \bar{\mathsf{x}} = 0$ translates into $\overline{w}_2 = 0$; further, from (b) we know also that $\overline{w}_1 = \mathsf{v}_1^\top \bar{\mathsf{x}} \neq 0$, and suppose that $\overline{w}_1 > 0$ (the opposite case $\overline{w}_1 < 0$ can be treated in complete symmetry). Next we fix all coordinates $w_j = \overline{w}_j$ for $j \geq 3$, so that $w_1^2 + w_2^2 = \overline{w}_1^2 + \overline{w}_2^2 = \overline{w}_1^2$, and thus $w_1 = +\sqrt{\overline{w}_1^2 - w_2^2}$ for all such w close to $\overline{\mathsf{w}}$. We conclude, removing now constant terms, that $\overline{w}_2 = 0$ is a local (unconstrained) minimizer of

$$a(w_2) = \frac{\lambda_1}{2}(\overline{w}_1^2 - w_2^2) + \frac{\lambda_2}{2} w_2^2 + \gamma_1 b(w_2) + \gamma_2 w_2 \,,$$

where $b(w_2) = \overline{w}_1 c(\frac{w_2}{\overline{w}_1})$ with $c(t) = \sqrt{1 - t^2}$. Note that the k-th derivative of b satisfies $b^{(k)}(w_2) = \overline{w}_1^{1-k} c^{(k)}(\frac{w_2}{\overline{w}_1})$. We calculate the first four derivatives:

$$\dot{c}(t) = -\frac{t}{c(t)} \,, \quad \ddot{c}(t) = -\frac{1}{c^3(t)} \,, \quad \dddot{c}(t) = -\frac{3t}{c^5(t)} \,, \quad c^{(4)}(0) = -3 \,.$$

Hence $b(0) = \overline{w}_1$ and

$$\dot{b}(0) = 0 \,, \quad \ddot{b}(0) = -\frac{1}{\overline{w}_1} \,, \quad \dddot{b}(0) = 0 \,, \quad b^{(4)}(0) = -\frac{3}{\overline{w}_1^3} \,.$$

Therefore, since $\gamma_2 = \mathsf{v}_2^\top \bar{\mathsf{x}} = 0$ by assumption (and the source of all this trouble), and since

$$0 = \mathsf{v}_1^\top \mathsf{o} = \mathsf{v}_1^\top [\mathsf{H}_{\bar{u}} \bar{\mathsf{x}} + \mathsf{q}] = \bar{\mathsf{x}}^\top \mathsf{H}_{\bar{u}} \mathsf{v}_1 + \mathsf{v}_1^\top \mathsf{q} = \bar{\mathsf{x}}^\top (\lambda_1 - \lambda_2) \mathsf{v}_1 + \gamma_1 = (\lambda_1 - \lambda_2) \overline{w}_1 + \gamma_1 \,,$$

we arrive at the conclusion that $\dot{a}(0) = \gamma_2 = 0$ as well as

$$\ddot{a}(0) = (\lambda_2 - \lambda_1) - \frac{\gamma_1}{\overline{w}_1} = 0$$

and $\dddot{a}(0) = \gamma_1 \dddot{b}(0) = 0$ as well, but finally

$$a^{(4)}(0) = -\frac{3\gamma_1}{\overline{w}_1^3} < 0\,,$$

which would contradict the fact that $\overline{w}_2 = 0$ is a local minimizer of a, so that $\bar{\mathsf{x}}$ cannot be a LNG of (13), contradicting the assumption $\gamma_2 = 0$.

(d) Now assume that $\bar{u} = -\lambda_2$. For the same $\mathsf{v} = \mathsf{v}_2 - \frac{\mathsf{v}_2^\top \bar{\mathsf{x}}}{\mathsf{v}_1^\top \bar{\mathsf{x}}}\mathsf{v}_1 \perp \bar{\mathsf{x}}$ as in (a) above we now arrive, using $\mathsf{H}_{\bar{u}}\mathsf{v}_2 = (\mathsf{Q} - \lambda_2 I)\mathsf{v}_2 = \mathsf{o}$ and (a) again, at the contradiction

$$0 \le \mathsf{v}^\top \mathsf{H}_{\bar{u}}\mathsf{v} = \alpha^2 \mathsf{v}_1^\top \mathsf{H}_{\bar{u}}\mathsf{v}_1 + (\mathsf{v}_2 + 2\alpha \mathsf{v}_1)^\top \mathsf{H}_{\bar{u}}\mathsf{v}_2 = \alpha^2(\lambda_1 - \lambda_2) < 0\,,$$

because $\alpha = -\frac{\mathsf{v}_2^\top \bar{\mathsf{x}}}{\mathsf{v}_1^\top \bar{\mathsf{x}}} \neq 0$ due to (c). Hence even $\bar{u} > -\lambda_2$. We conclude that $\mathsf{H}_{\bar{u}} = \mathsf{Q} + \bar{u}\mathsf{I}_n$ is nonsingular, otherwise we would obtain another eigenvalue of Q strictly between λ_1 and λ_2. □

3.3 The secular function; at most one LNG exists

For any LNG solution $\bar{\mathsf{x}}$ we have for the unique multiplier $\bar{u}$ that $\mathsf{H}_{\bar{u}}$ is nonsingular. The KKT conditions read therefore

$$\mathsf{H}_{\bar{u}}\bar{\mathsf{x}} = -\mathsf{q} \quad \text{or} \quad \bar{\mathsf{x}} = -(\mathsf{Q} + \bar{u}\mathsf{I}_n)^{-1}\mathsf{q}\,.$$

This implies that the value of the multiplier $\bar{u}$ in turn uniquely determines the KKT point $\bar{\mathsf{x}}$. Given the data (Q, q), we consider the *secular function*

$$\psi(u) := \left\|(\mathsf{Q} + u\mathsf{I}_n)^{-1}\mathsf{q}\right\|^2, \quad u \in \mathbb{R} \setminus \{-\lambda_n, \dots, -\lambda_1\}\,.$$

From Corollary 3.2(d) we conclude that $\bar{u}$ belongs to the domain of ψ and moreover is a 1-root of ψ, i.e. $\psi(\bar{u}) = \|\bar{\mathsf{x}}\|^2 = 1$. Also this function can be simplified by means of diagonalization of $\mathsf{Q} = \mathsf{V}\text{diag}\,(\lambda_1, \dots, \lambda_n)\mathsf{V}^\top$ with $\mathsf{V} = [\mathsf{v}_1, \dots, \mathsf{v}_n]$ being an orthonormal $n \times n$ matrix. Putting

$$\mu_i(u) := \frac{1}{\lambda_i + u} \neq 0 \quad (\text{in fact, } \mu_1(\bar{u}) < 0 < \mu_i(\bar{u}) \text{ for all } i \in \{2, \dots, n\}\,)\,,$$

we obtain $(\mathsf{Q} + u\mathsf{I}_n)^{-1} = \mathsf{V}\text{diag}\,(\mu_1(u), \dots, \mu_n(u))\mathsf{V}^\top$, then $(\mathsf{Q} + u\mathsf{I}_n)^{-1}\mathsf{q} = \sum_i \gamma_i \mu_i(u)\mathsf{v}_i$ with $\gamma_i = \mathsf{v}_i^\top \mathsf{q}$. We note for further use that

$$\bar{\mathsf{x}} = -\sum_{i=1}^{n} \gamma_i \mu_i(\bar{u})\mathsf{v}_i\,. \tag{14}$$

Finally, by orthonormality of all v_i,

$$\psi(u) = \sum_i \gamma_i^2 \mu_i^2(u) \geq 0 \quad \text{for all } u \in \mathbb{R} \setminus \{-\lambda_n, \ldots, -\lambda_1\} \, . \tag{15}$$

We calculate $\dot{\mu}_i(u) = -\mu_i^2(u)$ and thus

$$\begin{array}{rcccl} \dot{\psi}(u) & = & \sum_i \gamma_i^2 [\mu_i^2(u)]^{\cdot} & = & -2\sum_i \gamma_i^2 \mu_i^3(u)\,, \\ \ddot{\psi}(u) & = & -2\sum_i \gamma_i^2 [\mu_i^3(u)]^{\cdot} & = & 6\sum_i \gamma_i^2 \mu_i^4(u) > 0\,, \end{array}$$

and hence ψ is strictly convex on the open interval defined by $-\lambda_2 < u < -\lambda_1$. Hence there are at most two different 1-roots of ψ in this interval, and among these at most one u with $\dot{\psi}(u) \geq 0$. With some effort it can be shown that whenever $\bar{\mathsf{x}}$ is a LNG for (13) with multiplier $\bar{u}$, then indeed $\dot{\psi}(\bar{u}) \geq 0$. This way we arrive at Martínez' theorem [9]:

Theorem 3.1 *There is at most one LNG solution to (13).*

Proof. Following again [9], we define $\mathsf{w} = [\gamma_2\mu_2(\bar{u}), \ldots, \gamma_n\mu_n(\bar{u})]^\top \in \mathbb{R}^{n-1}$ and the $n \times (n-1)$ matrices

$$\mathsf{T} = \begin{bmatrix} \mathsf{w}^\top \\ -\gamma_1\mu_1(\bar{u})\mathsf{I}_{n-1} \end{bmatrix} \quad \text{as well as} \quad \mathsf{W} = \mathsf{VT}\,.$$

From Corollary 3.2(c) we infer $\gamma_1 = \mathsf{v}_1^\top \mathsf{q} \neq 0$, thus rank $\mathsf{W} = n-1$, and moreover $\mathsf{We}_j = \mathsf{VTe}_j = w_j \mathsf{v}_1 - \gamma_1\mu_1(\bar{u})\mathsf{v}_{j+1}$ which implies via (14)

$$\begin{array}{rcl} \bar{\mathsf{x}}^\top \mathsf{We}_j & = & -\sum\limits_{i=1}^n \gamma_i\mu_i(\bar{u})\mathsf{v}_i^\top [\gamma_{j+1}\mu_{j+1}(\bar{u})\mathsf{v}_1 - \gamma_1\mu_1(\bar{u})\mathsf{v}_{j+1}] \\ & = & \sum\limits_{i=1}^n \gamma_1\gamma_i\, \mu_1(\bar{u})\mu_i(\bar{u})\, \mathsf{v}_i^\top \mathsf{v}_{j+1} - \sum\limits_{i=1}^n \gamma_i\gamma_{j+1}\, \mu_i(\bar{u})\mu_{j+1}(\bar{u})\, \mathsf{v}_i^\top \mathsf{v}_1 \\ & = & \gamma_1\gamma_{j+1}\, \mu_1(\bar{u})\mu_{j+1}(\bar{u}) - \gamma_1\gamma_{j+1}\, \mu_1(\bar{u})\mu_{j+1}(\bar{u}) = 0 \end{array}$$

for all $j \in \{1, \ldots, n-1\}$, so that the columns of W form a basis for the hyperplane $\bar{\mathsf{x}}^\perp$. Therefore, by Corollary 3.2(a), the $(n-1) \times (n-1)$ matrix $\mathsf{B} := \mathsf{W}^\top \mathsf{H}_{\bar{u}} \mathsf{W}$ is positive-semidefinite, so that det $\mathsf{B} \geq 0$. Now

$$\mathsf{B} = \mathsf{T}^\top \mathsf{V}^\top \mathsf{QVT} + \bar{u}\mathsf{T}^\top \mathsf{T} = \mathsf{T}^\top \mathsf{DT}$$

with $\mathsf{D} = \mathrm{diag}\,(\lambda_1 + \bar{u}, \ldots, \lambda_n + \bar{u})$ (note that the upper left entry of D is negative). We will further rephrase B now and calculate det B then. First note that

$$\mathsf{DT} = \begin{bmatrix} (\lambda_1 + \bar{u})\mathsf{w}^\top \\ -\gamma_1\mu_1(\bar{u})\mathrm{diag}\,(\lambda_2 + \bar{u}, \ldots, \lambda_n + \bar{u}) \end{bmatrix} \text{and recall } \mathsf{T} = \begin{bmatrix} \mathsf{w}^\top \\ -\gamma_1\mu_1(\bar{u})\mathsf{I}_{n-1} \end{bmatrix}.$$

Hence for $\widehat{\mathsf{D}} = \gamma_1^2 \mu_1(\bar{u})^2 \mathrm{diag}\,(\lambda_2 + \bar{u}, \ldots, \lambda_n + \bar{u})$ we get

$$\mathsf{B} = \mathsf{T}^\top \mathsf{D} \mathsf{T} = (\lambda_1 + \bar{u}) \mathsf{w}\mathsf{w}^\top + \widehat{\mathsf{D}} = \widehat{\mathsf{D}} \left[\mathsf{I}_{n-1} + (\lambda_1 + \bar{u}) \widehat{\mathsf{D}}^{-1} \mathsf{w}\mathsf{w}^\top \right] ,$$

so that, using the formula $\det\,[\mathsf{I}_{n-1} + \mathsf{a}\mathsf{b}^\top] = 1 + \mathsf{b}^\top \mathsf{a}$, we arrive at

$$0 \leq \det\,\mathsf{B} = \det\,\widehat{\mathsf{D}} \left[1 + (\lambda_1 + \bar{u}) \mathsf{w}^\top \widehat{\mathsf{D}}^{-1} \mathsf{w} \right] . \tag{16}$$

Since $\det\,\widehat{\mathsf{D}} = [\gamma_1 \mu_1(\bar{u})]^{2n-2} \prod_{j=2}^n (\lambda_j + \bar{u}) > 0$, we deduce

$$1 + \mu_1^{-1}(\bar{u}) \mathsf{w}^\top \widehat{\mathsf{D}}^{-1} \mathsf{w} \geq 0 \,. \tag{17}$$

Now $[\gamma_1 \mu_1(\bar{u})]^2 \mathsf{w}^\top \widehat{\mathsf{D}}^{-1} \mathsf{w} = \mathsf{w}^\top \mathrm{diag}\,(\mu_2(\bar{u}), \ldots, \mu_n(\bar{u}))\, \mathsf{w} = \sum\limits_{j=2}^n \gamma_j^2 \mu_j^3(\bar{u})$, so that $[\mu_1(\bar{u})]^{-1} \mathsf{w}^\top \widehat{\mathsf{D}}^{-1} \mathsf{w} = \gamma_1^{-2} [\mu_1(\bar{u})]^{-3} \sum\limits_{j=2}^n \gamma_j^2 \mu_j^3(\bar{u})$, and (17) reduces to

$$1 + \frac{1}{\gamma_1^2 \mu_1^3(\bar{u})} \sum_{j=2}^n \gamma_j^2 \mu_j^3(\bar{u}) \geq 0$$

or, multiplying by $-2\gamma_1^2 \mu_1^3(\bar{u}) > 0$,

$$\dot{\psi}(\bar{u}) = -2 \sum_{j=1}^n \gamma_j^2 \mu_j^3(\bar{u}) \geq 0 \,,$$

which proves the assertion. Indeed, we showed there can be at most one multiplier $\bar{u}$ belonging to a LNG, and since $\bar{u}$ determines the LNG uniquely, there can be at most one LNG to (13). □

4 Duality of all-quadratic problems and SDPs

In this section we show that the Langrangian (and Wolfe) dual of an all-quadratic problem coincides with the dual of its SDP relaxation under mild assumptions. Also these results are well known and are discussed here only to complement above illustrations and counterexamples. Consider

$$\begin{array}{llll} q_0(\mathsf{x}) & \to & \min\ ! & \text{subject to} \\ q_i(\mathsf{x}) & \leq & 0\,, & i \in \{1, \ldots, m\}\,, \end{array} \tag{18}$$

where $q_i(\mathsf{x}) = \mathsf{x}^\top \mathsf{Q}_i \mathsf{x} - 2\mathsf{b}_i^\top \mathsf{x} + c_i$ are all quadratic functions (we may assume $c_0 = 0$) for $i \in \{0, 1, \ldots, m\}$. We denote by z_Q^* the optimal value (possibly not attained, or also equal to $-\infty$ (in the unbounded case) or to $+\infty$ (in the infeasible case).

We start by recalling a basic result on Schur complementation. To this end, recall the notion of the Moore-Penrose generalized inverse H^+ of a symmetric matrix H, which is again symmetric, of the same order as H, and satisfies $\mathsf{H}\mathsf{H}^+\mathsf{H} = \mathsf{H}$ as well as $\mathsf{H}^+\mathsf{H}\mathsf{H}^+ = \mathsf{H}^+$. A linear system $\mathsf{H}\mathsf{x} = \mathsf{d}$ is solvable in x if and only if $\mathsf{H}\mathsf{H}^+\mathsf{d} = \mathsf{d}$, in which case $\mathsf{x} = \mathsf{H}^+\mathsf{d}$ is the solution with the least distance to the origin o (if $\mathsf{H}\mathsf{x} = \mathsf{d}$ is inconsistent, then $\mathsf{H}\mathsf{H}^+\mathsf{d}$ is the least squares solution to this system). We abbreviate the fact that H is positive-semidefinite by $\mathsf{H} \succeq \mathsf{O}$. It is well known that $\mathsf{H} \succeq \mathsf{O}$ also implies $\mathsf{H}^+ \succeq \mathsf{O}$.

Lemma 4.1 *Let H be a symmetric $n \times n$ matrix, $\mathsf{v} \in \mathbb{R}^n$, and $\alpha \in \mathbb{R}$, and form the symmetric $(n+1) \times (n+1)$ matrix*

$$\mathsf{M} = \begin{bmatrix} \alpha & \mathsf{v}^\top \\ \mathsf{v} & \mathsf{H} \end{bmatrix} .$$

Then M is positive-semidefinite if and only if the following three conditions hold:

(a) H is positive-semidefinite;

(b) $\mathsf{v} \in \mathsf{H}(\mathbb{R}^n)$, or, equivalently, $\mathsf{v} \perp$ ker H;

(c) $\alpha \geq \mathsf{v}^\top \mathsf{H}^+ \mathsf{v}$.

Proof. Let M be positive-semidefinite. To establish (a), choose $\mathsf{x} = [0, \mathsf{y}^\top]^\top$ with $\mathsf{y} \in \mathbb{R}^n$ arbitrary; then $0 \leq \mathsf{x}^\top \mathsf{M} \mathsf{x} = \mathsf{y}^\top \mathsf{H} \mathsf{y}$. For (b), assume the contrary, i.e., that there is a vector $\mathsf{y} \in \ker \mathsf{H}$ such that $\mathsf{v}^\top \mathsf{y} < 0$, and choose, for large $t > 0$, the vector $\mathsf{x}_t = [1, t\mathsf{y}^\top]^\top$. Then $0 \leq \mathsf{x}_t^\top \mathsf{M} \mathsf{x}_t = \alpha + 2t\mathsf{v}^\top \mathsf{y} + 0 < 0$ if t is large enough, a contradiction. To show (c), take $\mathsf{x} = [-1, (\mathsf{H}^+\mathsf{v})^\top]^\top$ which gives

$$0 \leq \mathsf{x}^\top \mathsf{M} \mathsf{x} = \alpha - 2\mathsf{v}^\top \mathsf{H}^+ \mathsf{v} + \mathsf{v}^\top \mathsf{H}^+ \mathsf{H} \mathsf{H}^+ \mathsf{v} = \alpha - \mathsf{v}^\top \mathsf{H}^+ \mathsf{v} .$$

For sufficiency, we only need $\mathsf{v} = \mathsf{H}\mathsf{w}$ for some $\mathsf{w} \in \mathbb{R}^n$ and note that

$$\begin{bmatrix} \alpha & \mathsf{v}^\top \\ \mathsf{v} & \mathsf{H} \end{bmatrix} = (\alpha - \mathsf{v}^\top \mathsf{H}^+ \mathsf{v}) \begin{bmatrix} 1 \\ \mathsf{o} \end{bmatrix} [1, \mathsf{o}^\top] + \begin{bmatrix} \mathsf{w}^\top \\ \mathsf{I}_n \end{bmatrix} \mathsf{H} [\mathsf{w}\, \mathsf{I}_n] \succeq \mathsf{O} ,$$

because $\mathsf{v}^\top \mathsf{H}^+ \mathsf{v} = \mathsf{w}^\top \mathsf{H} \mathsf{H}^+ \mathsf{H} \mathsf{w} = \mathsf{w}^\top \mathsf{H} \mathsf{w}$. Thus the result. □

The counterexample $\mathsf{H} = 0$ in $\mathsf{M} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ shows that condition (b) above cannot be dispensed with.

Now we return to (18) and denote by $\mathsf{H_u} = \mathsf{Q}_0 + \sum_{i=1}^{m} u_i \mathsf{Q}_i$, by $\mathsf{d_u} = \mathsf{b}_0 + \sum_{i=1}^{m} u_i \mathsf{b}_i$ and by $\mathsf{c} = [c_1, \ldots, c_m]^\top$. Then the Lagrangian function and its derivatives w.r.t. x read

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

By the Frank/Wolfe theorem in its unrestricted (therefore easy) version, $\Theta(\mathsf{u}) := \inf\{L(\mathsf{x};\mathsf{u}) : \mathsf{x} \in \mathbb{R}^n\} > -\infty$ if and only if (a) $\mathsf{H_u}$ is positive-semidefinite; and (b) the linear equation system $\mathsf{H_u x} = \mathsf{d_u}$ has a solution. In this case, we have $\Theta(\mathsf{u}) = L(\mathsf{x};\mathsf{u})$ for any x with $\mathsf{H_u x} = \mathsf{d_u}$, in particular, say, for the least-norm solution $\mathsf{x_u} := \mathsf{H_u^+ d_u}$. So the Wolfe dual and the Lagrangian dual problem coincide, namely to

$$z^*_{DQ} := \sup\left\{L(\mathsf{x_u};\mathsf{u}) : \mathsf{H_u} \succeq \mathsf{O}\,,\ \mathsf{H_u H_u^+ d_u} = \mathsf{d_u}\right\}\,, \tag{19}$$

where $\mathsf{H_u} \succeq \mathsf{O}$ denotes positive-semidefiniteness of $\mathsf{H_u}$ and the other condition exactly characterizes solvability of the system $\mathsf{H_u x} = \mathsf{d_u}$ in x. Weak duality ensures $z^*_{DQ} \leq z^*_Q$. Furthermore, by (19), for any $(\mathsf{x},\mathsf{u}) \in \mathbb{R}^n \times \mathbb{R}^m_+$ with $\mathsf{H_u} \succeq 0$ and $\mathsf{H_u x} = \mathsf{d_u}$, (in particular, for $\mathsf{x} = \mathsf{x_u} = \mathsf{H_u^+ d_u}$ in case $\mathsf{H_u H_u^+ d_u} = \mathsf{d_u}$) we see

$$\Theta(\mathsf{u}) = L(\mathsf{x};\mathsf{u}) = \mathsf{x}^\top \mathsf{d}_u - 2\mathsf{d_u}^\top \mathsf{x} + \mathsf{c}^\top \mathsf{u} = \mathsf{c}^\top \mathsf{u} - \mathsf{d_u}^\top \mathsf{x}\,. \tag{20}$$

We pass to the semidefinite relaxation of the problem (18): to this end we need the symmetric $(n+1)\times(n+1)$ matrices

$$\mathsf{M}_i = \begin{bmatrix} c_i & -\mathsf{b}_i^\top \\ -\mathsf{b}_i & \mathsf{Q}_i \end{bmatrix}\,, \quad i \in \{0,1,\ldots,m\}\,.$$

The SDP relaxation now uses Frobenius duality $\langle \mathsf{X},\mathsf{S}\rangle = \mathrm{trace}(\mathsf{XS})$ on matrices of this order and reads in its primal form

$$z^*_{SP} := \inf\left\{\langle \mathsf{M}_0,\mathsf{X}\rangle : \langle \mathsf{M}_i,\mathsf{X}\rangle \leq 0\,,\ \langle \mathsf{J}_0,\mathsf{X}\rangle = 1\,,\ \mathsf{X} \succeq \mathsf{O}\right\} \tag{21}$$

with $\mathsf{J}_0 = [1,0,\ldots,0]^\top [1,0,\ldots,0]$, while its dual is given by

$$z^*_{SD} := \sup\left\{y_0 \in \mathbb{R} : \mathsf{Z}(\mathsf{y}) \succeq \mathsf{O}, \mathsf{y} = [y_0, \mathsf{u}^\top]^\top \in \mathbb{R}\times\mathbb{R}^m_+\right\}\,, \tag{22}$$

where

$$\mathsf{Z}(\mathsf{y}) := \mathsf{M}_0 - y_0 \mathsf{J}_0 + \sum_{i=1}^{m} u_i \mathsf{M}_i = \begin{bmatrix} \mathsf{c}^\top \mathsf{u} - y_0 & -\mathsf{d_u}^\top \\ -\mathsf{d_u} & \mathsf{H_u} \end{bmatrix} \tag{23}$$

is the slack matrix.

The standard lifting $\mathsf{X} = [1, \mathsf{x}^\top]^\top [1, \mathsf{x}^\top]$ shows $z^*_{SP} \leq z^*_Q$, and weak duality of the primal-dual SDP pair shows $z^*_{SD} \leq z^*_{SP}$. It can easily be shown that strict feasibility of (18) implies strict feasibility of (21). Moreover, if Q_i is (strictly) positive-definite for at least one $i \in \{1, \ldots, m\}$, then also (22) is strictly feasible, so that full strong duality holds for the primal-dual SDP pair: $z^*_{SD} = z^*_{SP}$, this optimal value is attained in both SPDs, i.e., there is an $\mathsf{X}^* \succeq \mathsf{O}$ feasible to (21) and a $\mathsf{y}^* = [y^*_0, (\mathsf{u}^*)^\top]^\top \in \mathbb{R} \times \mathbb{R}^m_+$ feasible to (22) such that $\langle \mathsf{M}_0, \mathsf{X}^* \rangle = z^*_{SP} = z^*_{SD} = y^*_0$ and, by complementary slackness of the positive-semidefinite matrices, $\mathsf{Z}(\mathsf{y}^*)\mathsf{X}^* = \mathsf{O}$; in particular, the first column z of $\mathsf{Z}(\mathsf{y}^*)\mathsf{X}^*$ must equal the zero vector. Now decompose the symmetric $(n+1) \times (n+1)$ matrix X^*

$$\mathsf{X}^* = \begin{bmatrix} 1 & (\mathsf{x}^*)^\top \\ \mathsf{x}^* & \mathsf{Y}^* \end{bmatrix} \quad \text{so that} \quad \begin{bmatrix} \mathsf{c}^\top \mathsf{u}^* - y^*_0 - \mathsf{d}^\top_{\mathsf{u}^*} \mathsf{x}^* \\ -\mathsf{d}_{\mathsf{u}^*} + \mathsf{H}_{\mathsf{u}^*} \mathsf{x}^* \end{bmatrix} = \mathsf{z} = \begin{bmatrix} 0 \\ \mathsf{o} \end{bmatrix}, \tag{24}$$

due to (23); in particular, we get $\mathsf{H}_{\mathsf{u}^*}\mathsf{x}^* = \mathsf{d}_{\mathsf{u}^*}$ and, by $\mathsf{Z}(\mathsf{y}^*) \succeq \mathsf{O}$, also $\mathsf{H}_{\mathsf{u}^*} \succeq \mathsf{O}$. Hence, by (20) and (24),

$$z^*_{DQ} \geq \Theta(\mathsf{u}^*) = \mathsf{c}^\top \mathsf{u}^* - \mathsf{d}^\top_{\mathsf{u}^*} \mathsf{x}^* = y^*_0 = z^*_{SD} \,. \tag{25}$$

We will show that in fact equality $z^*_{SD} = z^*_{DQ}$ holds under these conditions. To this end, consider an arbitrary $\mathsf{u} \in \mathbb{R}^m_+$ feasible to (19) and define $y_0 = \Theta(\mathsf{u}) = \mathsf{c}^\top \mathsf{u} - \mathsf{d}^\top_{\mathsf{u}} \mathsf{x}_{\mathsf{u}} = \mathsf{c}^\top \mathsf{u} - \mathsf{d}^\top_u \mathsf{H}^+_{\mathsf{u}} \mathsf{d}_{\mathsf{u}}$; see again (20). By Lemma 4.1, it follows from (23) that $\mathsf{Z}(\mathsf{y}) \succeq \mathsf{O}$ for $\mathsf{y} = [y_0, \mathsf{u}^\top]^\top \in \mathbb{R} \times \mathbb{R}^m_+$ so that the latter vector is feasible to (22). Hence $\Theta(\mathsf{u}) = y_0 \leq z^*_{SD}$ and therefore the reverse inequality follows; summarizing, we obtain, under strict feasibility of (18) and strict positive-definiteness of at least one Q_i, $1 \leq i \leq m$, that

$$z^*_{DQ} = z^*_{SD} = z^*_{SP} \leq z^*_Q \,.$$

Remark 4.1 *Continuing with the example of Remark 1.3, we see that, at the (non-isolated) global solution* $\mathsf{x}^* = \mathsf{o}$ *and for any* $\mathsf{u} \in \mathbb{R}^2_+ \setminus \{\mathsf{u}^*\}$*, the Hessian* H_{u} *is indefinite so* $\Theta(\mathsf{u}) = -\infty$ *for all* $\mathsf{u} \neq \mathsf{u}^* = [1, 0]^\top$*. On the other hand, obviously* $\Theta(\mathsf{u}^*) = 0$*. So* $z^*_{DQ} = 0 = z_Q*$ *and* $(\mathsf{x}^*, \mathsf{u}^*)$ *is a primal-dual optimal pair, so full strong duality holds. Moreover, we concluded above that* $z^*_{SD} \geq z^*_{DQ}$ *is always true even without any strict feasibility, so in fact we get, by weak duality for the SDP and the fact that this is a relaxation of the original problem*

$$0 = z^*_{DQ} \leq z^*_{SD} \leq z^*_{SP} \leq z^*_Q = 0 \,,$$

despite the fact that

$$\mathsf{Z}(\mathsf{y}) = \begin{bmatrix} -y_0 & \mathsf{o}^\top \\ \mathsf{o} & \mathsf{H}_{\mathsf{u}} \end{bmatrix}$$

can never be positive-definite (not even for $\mathsf{u} = \mathsf{u}^* = [1, 0]^\top$*). Needless to stress that no* Q_i *is positive-definite in this example.*

Acknowledgements. The author is indebted to the Isaac Newton Institute at Cambridge University where this paper was completed during his participation as a visiting fellow in the Polynomial Optimization Programme 2013, organized by Joerg Fliege, Jean-Bernard Lasserre, Adam Letchford and Markus Schweighofer, for providing a stimulating atmosphere and interesting events. I also am indebted to Andreas Fischer, Michael Overton and in particular to Oliver Stein for valuable suggestions and hints. James V. Burke's site http://www.math.washington.edu/~burke/crs/408/ on Nonlinear Optimization is warmly recommended.

References

[1] Immanuel M. Bomze. Copositivity conditions for global optimality in indefinite quadratic programming problems. *Czechosl. J. Operations Res.*, 1:7–19, 1992.

[2] Immanuel M. Bomze, Vladimir Demyanov, Roger Fletcher, Tamás Terlaky, and Imre Pólik. *Nonlinear Optimization*, volume 1989 of *Lecture Notes in Mathematics*. Springer-Verlag, New York, 2010.

[3] Jonathan M. Borwein. Necessary and sufficient conditions for quadratic minimality. *Numer. Funct. Anal. Optim.*, 5:127–140, 1982.

[4] Luis B. Contesse. Une caractérisation complète des minima locaux. *Numer. Math.*, 34:315–332, 1980.

[5] Gabriele Danninger. Role of copositivity in optimality criteria for nonconvex optimization problems. *J. Optim. Theory Appl.*, 75(3):535–558, 1992.

[6] Ulrich Faigle, Walter Kern, and Georg Still. *Algorithmic Principles of Mathematical Programming.* Kluwer Texts in the Mathematical Sciences Vol.24. Springer, Dordrecht, 2002.

[7] Alexey Feridovich Izmailov and Mikhail V. Solodov. Examples of dual behaviour of Newton-type methods on optimization problems with degenerate constraints. *Comput. Optimiz. Appl.*, 42:231–264, 2009.

[8] Antal Majthay. Optimality conditions for quadratic programming. *Math. Programming*, 1:359–365, 1971.

[9] José Mario Martínez. Local minimizers of quadratic functions on euclidean balls and spheres. *SIAM J. Optim.*, 4(1):159–176, 1994.

[10] David W. Peterson. A review of constraint qualifications in finite-dimensional spaces. *SIAM Rev.*, 15(3):639–654, 1973.