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 linearization 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. Application to the trust-region problem and several (counter-)examples illustrate the advantage of this approach.

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

February 19, 2015

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

1.1 A new constraint qualification

To begin with, let us shortly discuss several *constraint qualifications (CQ)*, most of them well studied but one apparently novel, for a smooth optimization problem

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

which can be written in the more compact form

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

$h(\mathsf{x}) = [h_1(\mathsf{x}), \ldots h_q(\mathsf{x})]^\top \in \mathbb{R}^q$ and $g(\mathsf{x}) = [g_1(\mathsf{x}), \ldots, g_m(\mathsf{x})]^\top \in \mathbb{R}^m$. 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. the scalar variable t). As usual, for any cone $C \subseteq \mathbb{R}^n$, we denote its dual (or polar) cone by

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

This notation should not be confused with C^* used for the dual space of C (which is not needed here). In fact, we have $C^\star \cap (-C)^\star = C^\perp$, which will be used later on.

Definition 1.1 *Let $\mathsf{x} \in M$ be a feasible point of problem (1) and denote by $I(\mathsf{x}) = \{i \in \{1,\ldots,m\} : g_i(\mathsf{x}) = 0\}$ 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 linearization cone *of M at x. Later on, we will also use the* reduced linearization cone *of directions in $\Gamma(\mathsf{x})$ along which no first-order change in f happens:*

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

Finally, consider the (derivative) tangent (or contingent in Bouligand's sense) cone

$$T_M(\mathsf{x}) = \left\{\mathsf{v} \in \mathbb{R}^n : \mathsf{v} = \lim_{k\to\infty} \mathsf{v}_k \textit{ with } \mathsf{x} + t_k \mathsf{v}_k \in M,\ \textit{some } t_k \searrow 0 \textit{ as } k \to \infty\right\}\ .$$

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 not be polyhedral, not even convex).

Definition 1.2 *Consider problem (1) with the cones from Definition 1.1.*

(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. [7, 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 $\mathsf{v} = \lim_{t \searrow 0} \mathsf{v}_t$ with $\mathsf{v}_t = \frac{1}{t}[\mathsf{y}(t) - \mathsf{y}(0)]$ and $\bar{\mathsf{x}} + t\mathsf{v}_t = \mathsf{y}(t) \in M$.*

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 an $\varepsilon > 0$ and 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 $0 \le t < \varepsilon$, as desired. $\square$

Most of the following relations between the CQs are well known; for a standard reference see [12].

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 $\mathsf{v} \in \Gamma(\bar{\mathsf{x}})$ 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 $i \in I(\bar{\mathsf{x}})$. This d exists by the MFCQ, and without loss of generality we may assume $\|\mathsf{d}\| = 1$. Clearly, for any $k \in \mathbb{N}$ also $\mathsf{d}_k := \mathsf{v} + \frac{1}{k}\mathsf{d} \in \Gamma(\bar{\mathsf{x}})$ satisfies $\mathsf{d}_k^\top \nabla g_i(\bar{\mathsf{x}}) < 0$ for all $i \in I(\bar{\mathsf{x}})$, so by Lemma 1.1 there is an $\varepsilon_k > 0$ and a trajectory $\{\mathsf{y}_k(t) : t \in [0, \varepsilon_k]\} \subseteq M$ starting from $\bar{\mathsf{x}}$, i.e. $\mathsf{y}_k(0) = \bar{\mathsf{x}}$, and having starting tangent $\dot{\mathsf{y}}_k(0) = \mathsf{d}_k$. So for arbitrary $\delta > 0$, choose $t_k \le \min\left\{\frac{1}{k}, t_{k-1}, \varepsilon_k\right\}$ so small that $\|\mathsf{d}_k - \mathsf{v}_k\| \le \frac{1}{k}$ for all $k \in \mathbb{N}$, where we put $\mathsf{v}_k := \frac{1}{t_k}[\mathsf{y}_k(t_k) - \bar{\mathsf{x}}]$. Then $\|\mathsf{v} - \mathsf{d}_k\| = \frac{1}{k} \le \frac{\delta}{2}$ and thus $\|\mathsf{v} - \mathsf{v}_k\| \le \delta$ if $k \ge \frac{2}{\delta}$, so $\mathsf{v}_k \to \mathsf{v}$ as $k \to \infty$ and $\bar{\mathsf{x}} + t_k\mathsf{v}_k = \mathsf{y}_k(t_k) \in M$

as well as $t_k \searrow 0$ as $k \to \infty$ by construction. So $\mathsf{v} \in T_M(\bar{\mathsf{x}})$ and hence the ACQ are met. Obviously, ACQ implies both RACQ and GCQ. $\square$

Remark 1.1 *This example is taken from James V. Burke's 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* $\mathbb{R}^2_+$.

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

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

more precisely, for the objective function f, the equality constraints h_i and $g_i(u_i > 0)$, and the inequality constraints with $g_i(u_i = 0)$. For a KKT point $\bar{\mathsf{x}}$ with multipliers u, the reduced linearization cone $\Gamma_0(\bar{\mathsf{x}})$ coincides with the linearization cone of M_u because

$$\mathsf{v} \perp \nabla f(\bar{\mathsf{x}}) = -\sum_{i=1}^m u_i \nabla g_i(\bar{\mathsf{x}}) - \sum_{i=1}^q u_{i+m} \nabla_i h_i(\bar{\mathsf{x}})$$

and $\mathsf{v} \in \Gamma(\bar{\mathsf{x}})$ implies $\mathsf{v} \perp \nabla g_i(\bar{\mathsf{x}})$ if $u_i > 0$. Then the RACQ for M_u reads

$$\Gamma_0(\bar{\mathsf{x}}) \subseteq T_{M_\mathsf{u}}(\mathsf{x}) \cup [-T_{M_\mathsf{u}}(\mathsf{x})] \,, \tag{2}$$

where the explicit description of M_u as detailed above only enters via $\Gamma_0(\bar{\mathsf{x}})$. Note that even if M is convex and satisfies all CQs (e.g. if M has a Slater point), the reduced feasible set M_u is typically non-convex and has empty interior, see, e.g. [3, Figure 1].

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 [6, Ex.2.3] shows:

Remark 1.2 *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 linearization cone is* $\Gamma(\bar{\mathsf{x}}) = \{\mathsf{v} \in \mathbb{R}^2 :$

$v_2 \geq 0\}$. *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*

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

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 from* $\bar{\mathsf{x}} = \mathsf{o}$. *But* $-\mathsf{v}$ *is such a tangent vector, so RACQ holds. Similarly, also for* $\mathsf{u} = [0,1]^\top$, *ACQ is not met by*

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

Here $\mathsf{v} = [-1,0]^\top$ *is not a starting tangent vector but* $-\mathsf{v}$ *is one, so again RACQ holds. All these examples also violate the GCQ.*

From Remarks 1.1 and 1.2, we see that the weakest CQ used in the usual context (e.g., to ensure existence of Lagrange multipliers at local solutions), namely GCQ, neither does imply, nor implies, the RACQ. Further, as noted by a diligent referee, the RACQ imply $\Gamma(\bar{\mathsf{x}}) \setminus T_M(\bar{\mathsf{x}}) \subseteq \{\nabla g_i(\bar{\mathsf{x}}) : i \in I(\bar{\mathsf{x}})\}^\perp$. This is not so strange as it may seem at first sight, as RACQ is used for *second-order* rather than for first-order optimality conditions like MFCQ, ACQ or GCQ. For instance, extending the CQ from ACQ to RACQ is decisive in a recent study of the famous Celis-Dennis-Tapia problem (optimizing a quadratic over the intersection of two ellipsoids) in [3].

1.2 Second-order conditions for local optimality

For problem (1), we define the *Lagrangian 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 $i \in \{1, \ldots, m\}$ and $u_i \in \mathbb{R}$ for all $i \in \{m+1, \ldots m+q\}$ 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 [4] who apparently had the final word up to now in a series of publications dealing with similar second-order optimality conditions (e.g., [5, 10]).

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

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

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 (2) at $\bar{\mathsf{x}}$, 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 $\mathsf{x}_k \in M \setminus \{\bar{\mathsf{x}}\}$ with $t_k := \|\mathsf{x}_k - \bar{\mathsf{x}}\| \searrow 0$ and $\rho_k \to 0$ as $k \to \infty$ with

$$f(\mathsf{x}_k) < f(\bar{\mathsf{x}}) + \rho_k t_k^2 \quad \text{as } k \to \infty\,. \tag{3}$$

Consider the directions $\mathsf{v}_k = \frac{1}{t_k}(\mathsf{x}_k - \bar{\mathsf{x}})$ of unit length and assume without loss of generality that $\mathsf{v}_k \to \mathsf{v}$ as $k \to \infty$. Then $\mathsf{v} \in \Gamma(\bar{\mathsf{x}})$ as noted after Definition 1.1. Obviously

$$\mathsf{v}^\top \nabla f(\bar{\mathsf{x}}) = \lim_{k\to\infty} \mathsf{v}_k^\top \nabla f(\bar{\mathsf{x}}) = \lim_{k\to\infty} \frac{1}{t_k}[f(\mathsf{x}_k) - f(\bar{\mathsf{x}})] \leq \lim_{k\to\infty} \frac{\rho_k t_k^2}{t_k} = 0\,,$$

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 (3),

$$\begin{array}{rcl} f(\bar{\mathsf{x}}) + \rho_k t_k^2 & > & f(\mathsf{x}_k) \geq L(\mathsf{x}_k;\bar{\mathsf{u}}) = L(\bar{\mathsf{x}};\bar{\mathsf{u}}) + \frac{t_k^2}{2}\, \mathsf{v}_k^\top D^2_{\mathsf{x}} L(\bar{\mathsf{x}};\bar{\mathsf{u}})\mathsf{v}_k + \mathsf{o}(t_k^2) \\ & = & f(\bar{\mathsf{x}}) + \frac{t_k^2}{2}\, \mathsf{v}_k^\top D^2_{\mathsf{x}} L(\bar{\mathsf{x}};\bar{\mathsf{u}})\mathsf{v}_k + \mathsf{o}(t_k^2)\,, \end{array}$$

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

$$\rho_k \geq \tfrac{1}{2}\, \mathsf{v}_k^\top D_{\mathsf{x}}^2 L(\bar{\mathsf{x}};\bar{\mathsf{u}})\mathsf{v}_k + \mathfrak{o}(1) \geq \tfrac{1}{3}\, \mathsf{v}^\top D_{\mathsf{x}}^2 L(\bar{\mathsf{x}};\bar{\mathsf{u}})\mathsf{v} > 0$$

for all large enough $k \in \mathbb{N}$, a contradiction.
(b) Suppose $\mathsf{v} \in \Gamma_0(\bar{\mathsf{x}})$ satisfies $\mathsf{v}^\top D_{\mathsf{x}}^2 L(\bar{\mathsf{x}};\bar{\mathsf{u}})\mathsf{v} < 0$. 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 $\mathsf{v}^\top D_{\mathsf{x}}^2 L(\bar{\mathsf{x}};\bar{\mathsf{u}})\mathsf{v}$. Then choose a close enough direction v_k and step sizes $t_k \searrow 0$ as $k \to \infty$ such that $\mathsf{x}_k = \bar{\mathsf{x}} + t_k \mathsf{v}_k \in M_{\bar{\mathsf{u}}}$. Then we have by continuity also

$$\mathsf{v}_k^\top D_{\mathsf{x}}^2 L(\bar{\mathsf{x}};\bar{\mathsf{u}})\mathsf{v}_k \leq \tfrac{1}{2}\, \mathsf{v}^\top D_{\mathsf{x}}^2 L(\bar{\mathsf{x}};\bar{\mathsf{u}})\mathsf{v} < 0$$

if $k \in \mathbb{N}$ is large enough, and therefore, using $f(\mathsf{x}) = L(\mathsf{x};\bar{\mathsf{u}})$ for all $\mathsf{x} \in M_{\bar{\mathsf{u}}}$,

$$\begin{aligned} f(\mathsf{x}_k) &= L(\mathsf{x}_k;\bar{\mathsf{u}}) = L(\bar{\mathsf{x}};\bar{\mathsf{u}}) + \tfrac{t_k^2}{2}\, \mathsf{v}_k^\top D_{\mathsf{x}}^2 L(\bar{\mathsf{x}};\bar{\mathsf{u}})\mathsf{v}_k + \mathfrak{o}(t_k^2) \\ &\leq f(\bar{\mathsf{x}}) + \tfrac{t_k^2}{4}\, \mathsf{v}^\top D_{\mathsf{x}}^2 L(\bar{\mathsf{x}};\bar{\mathsf{u}})\mathsf{v} + \mathfrak{o}(t_k^2) < f(\bar{\mathsf{x}}) \end{aligned}$$

if k is large enough, contradicting local optimality of $\bar{\mathsf{x}}$. □

Remark 1.3 *The necessary second-order conditions can also be satisfied even if RACQ fails: this example is due to F. Facchinei, A. Fischer (personal communication) and M. Herrich who adapted [9, Ex.2.2], see also the references therein; also cf. [6, 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* $\bar{\mathsf{x}} = \mathsf{o}$ *is optimal. Further, the Lagrangian has derivatives*

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

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* $\bar{\mathsf{u}} = [1,0]^\top$, *while at* $\bar{\mathsf{x}}$, *any* $\mathsf{u} \in \mathbb{R}_+^2$ *satisfy the KKT system* $\nabla_{\mathsf{x}} L(\bar{\mathsf{x}};\mathsf{u}) = \mathsf{o}$. *It is easy to see that there are no other KKT points for this problem. Next we investigate*

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

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

$$\Gamma(\mathsf{x}_t^\pm) = \left\{ \mathsf{v} \in \mathbb{R}^3 : t v_2 \leq \pm t v_3 \text{ and } \pm t v_1 \leq 0 \right\},$$

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

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

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

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

because for these choices of u *we have to use* g_2 *as inequality and* g_1 *as equality constraint. Note that* $M_{\bar{\mathsf{u}}}$ *is a closed cone satisfying* $-M_{\bar{\mathsf{u}}} = M_{\bar{\mathsf{u}}}$ *and*

$$[M_{\bar{\mathsf{u}}}]^{\perp} \subseteq \left\{[0,1,1]^{\top}, [0,-1,1]^{\top}, [1,0,-1]^{\top}\right\}^{\perp} = \{\mathsf{o}\}\,,$$

so

$$T_{M_{\bar{\mathsf{u}}}}(\bar{\mathsf{x}}) = T_{M_{\bar{\mathsf{u}}}}(\mathsf{o}) = M_{\bar{\mathsf{u}}}\,.$$

Hence $[T_{M_{\bar{\mathsf{u}}}}(\bar{\mathsf{x}})]^{\star} = [M_{\bar{\mathsf{u}}}]^{\star} = [M_{\bar{\mathsf{u}}}]^{\perp} = \{\mathsf{o}\} = [\Gamma_0(\bar{\mathsf{x}})]^{\star}$*, and GCQ is satisfied for* $M_{\bar{\mathsf{u}}}$ *at* $\bar{\mathsf{x}}$.

We further illustrate the necessity for a CQ like RACQ to enforce the second-order optimality condition by another example where there is a unique optimal solution (in above example $\bar{\mathsf{x}}$ is not an isolated (local) solution to the problem), involving a particular constellation of the Celis-Dennis-Tapia problem which was one motivation of this study.

Remark 1.4 *Suppose* $q = 0$*,* $m = 2$*, and that* $\{\mathsf{x} : g_2(\mathsf{x}) \le 0\} \subset \{x : g_1(\mathsf{x}) \le 0\}$*. Further suppose that* $g_1(\bar{\mathsf{x}}) = g_2(\bar{\mathsf{x}}) = 0$ *and that*

$$g_1(\mathsf{x}) = 0 \ge g_2(\mathsf{x}) \quad \Longrightarrow \quad \mathsf{x} = \bar{\mathsf{x}}$$

(think of a small disc touching inside a larger one at $\bar{\mathsf{x}}$*). Then* $\bar{\mathsf{x}}$ *is the only global solution to* $\min\{f(\mathsf{x}) := -g_1(\mathsf{x}) : g_i(\mathsf{x}) \le 0\,, i = 1,2\}$*, and by rescaling* g_2 *if needed, we may assume* $\bar{\mathsf{y}} := \nabla g_1(\bar{\mathsf{x}}) = \nabla g_2(\bar{\mathsf{x}})$*. Take the multiplier pair* $\bar{\mathsf{u}} = [1,0]^{\top}$ *(note that LICQ fails but MFCQ holds at* $\bar{\mathsf{x}}$*). Then*

$$M_{\bar{\mathsf{u}}} = \{x : g_1(\mathsf{x}) = 0 \ge g_2(\mathsf{x})\} = \{\bar{\mathsf{x}}\} \quad \text{and} \quad T_{M_{\bar{\mathsf{u}}}}(\bar{\mathsf{x}}) = \{\mathsf{o}\}\,,$$

but $\Gamma_0(\bar{\mathsf{x}}) = \bar{\mathsf{y}}^{\perp}$*, so RACQ and all other CQs in Definition (1.2) are violated. And indeed, the example given by [13, (2.35),(2.36)] exhibits a Hessian*

$$D^2_{\mathsf{x}} L(\bar{\mathsf{x}}; \bar{\mathsf{u}}) = \begin{bmatrix} -\beta & 0 \\ 0 & -\beta \end{bmatrix}$$

for a certain $\beta > 0$*, which cannot be* $\Gamma_0(\bar{\mathsf{x}})$*-copositive.*

1.3 Quadratic optimization over polyhedra (QP)

Theorem 1.1 says

$$\text{strict copositivity} \Rightarrow \text{strict local solution} \Rightarrow \text{local solution} \Rightarrow \text{copositivity}\,, \tag{4}$$

where we need RACQ (2) on $M_{\bar{\mathsf{u}}}$ for the rightmost implication. For quadratic optimization problems over polyhedra, the leftmost and the rightmost implications in (4) become equivalences (without any CQ, as usual with linear constraints). This has been known before, see, e.g. [2, p.5] and the references therein. But in the strict case, one can specify a distance error bound (implying even strong optimality rather than strict optimality), and this will done below (apparently for the first time in literature), along with an explicit construction of the involved constants.

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 $\delta := \min\limits_{\mathsf{v} \in B_0} \mathsf{v}^\top \mathsf{Q}\mathsf{v} > 0$. 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 $s < i \le k$. 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 $\mu_i \ge 0$ 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

$$\left.\begin{aligned} \|\mathsf{y}\| &\le \sum_{i>s} \mu_i \|\mathsf{r}_i - \bar{\mathsf{r}}_i\| = \sum_{i>s} \mu_i \sqrt{1 - (\mathsf{c}^\top \mathsf{r}_i)^2} \\ &\le \sum_{i>s} \mu_i \eta (\mathsf{c}^\top \mathsf{r}_i) = \eta \|\mathsf{w}\|\,, \end{aligned}\right\} \tag{5}$$

where $\eta = \max\limits_{s<i\le k} \frac{\sqrt{1-(\mathsf{c}^\top \mathsf{r}_i)^2}}{\mathsf{c}^\top \mathsf{r}_i}$. 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

$$\left.\begin{aligned} \|\mathsf{v}\| &\le (1+\eta)\|\mathsf{w}\| + \|\mathsf{z}\| \quad \text{and thus} \\ \|\mathsf{v}\|^2 &\le (1+\eta)(2+\eta)\|\mathsf{w}\|^2 + (2+\eta)\|\mathsf{z}\|^2\,. \end{aligned}\right\} \tag{6}$$

Next choose a number $\beta > 0$ such that $|\mathsf{p}^\top \mathsf{Q} \mathsf{q}| \le 2\beta \|\mathsf{p}\|\, \|\mathsf{q}\|$ 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 (5) that

$$\left.\begin{aligned} & \mathsf{c}^\top \mathsf{v} + \tfrac{1}{2} \mathsf{v}^\top \mathsf{Q} \mathsf{v} \\ = \;& \mathsf{c}^\top \mathsf{w} + \tfrac{1}{2} \mathsf{w}^\top \mathsf{Q} \mathsf{w} + \mathsf{w}^\top \mathsf{Q} \mathsf{y} + \mathsf{w}^\top \mathsf{Q} \mathsf{z} + \mathsf{y}^\top \mathsf{Q} \mathsf{z} + \tfrac{1}{2} \mathsf{y}^\top \mathsf{Q} \mathsf{y} + \tfrac{1}{2} \mathsf{z}^\top \mathsf{Q} \mathsf{z} \\ \ge \;& \|\mathsf{w}\| - \beta \|\mathsf{w}\|^2 - 2\beta \|\mathsf{w}\|\, (\|\mathsf{y}\| + \|\mathsf{z}\|) - 2\beta \|\mathsf{y}\|\, \|\mathsf{z}\| - \beta \|\mathsf{y}\|^2 + \delta \|\mathsf{z}\|^2 \\ \ge \;& \|\mathsf{w}\|\, \big(1 - \beta \|\mathsf{w}\| - 2\beta\eta \|\mathsf{w}\| - 2\beta \|\mathsf{z}\| - 2\beta\eta \|\mathsf{z}\| - \beta\eta^2 \|\mathsf{w}\|\big) + \delta \|\mathsf{z}\|^2 \\ = \;& \|\mathsf{w}\|\, \big(1 - \beta(1+\eta)^2 \|\mathsf{w}\| - 2\beta(1+\eta)\|\mathsf{z}\|\big) + \delta \|\mathsf{z}\|^2\,. \end{aligned}\right\} \tag{7}$$

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 $\|\mathsf{v}\| \le \varepsilon$ entails both $\|\mathsf{w}\| \le \varepsilon$ and also $\|\mathsf{y} + \mathsf{z}\| \le \varepsilon$. Therefore also

$$\|\mathsf{z}\| \le \|\mathsf{y} + \mathsf{z}\| + \|\mathsf{y}\| \le \varepsilon + \eta \|\mathsf{w}\| \le \varepsilon(1+\eta)\,. \tag{8}$$

Now the factor of $\|\mathsf{w}\|$ in the last line of (7) exceeds $(1+\eta)\delta\|\mathsf{w}\|$ if

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

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

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

using the relations $\|\mathsf{w}\| \le \varepsilon$ and $\|\mathsf{z}\| \le \varepsilon(1+\eta)$ established in (8). Then we arrive via (7) and (6) at

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

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) is proved in [1, Lemma 1]. The proof of (a) above is a refinement of the arguments there. □

Now we are in a position to state and prove local optimality criteria; part (b) of below theorem can be found in [1, 6], see also [2, Sect.1.1], but part (a) is new.

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

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

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. □

Remark 1.5 *As a diligent referee pointed out, the implication "(strict) copositivity $\Longrightarrow$ (strict) optimality" above holds for all quadratic f even M is not a polyhedron but just a convex set.*

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 as, e.g., in the Celis-Dennis-Tapia problem.

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*

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

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

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

and the assertion is shown. $\square$

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

$$\min\left\{f(\mathsf{x}) = \tfrac{1}{2}\,\mathsf{x}^\top \mathsf{Q}\mathsf{x} + \mathsf{q}^\top\mathsf{x} : \mathsf{x}\in M\right\}\,, \tag{9}$$

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

$$\Gamma(\bar{\mathsf{x}}) = \mathbb{R}_+(M-\bar{\mathsf{x}}) = \left\{\mathsf{v}\in\mathbb{R}^n : \mathsf{a}_i^\top\mathsf{v}\le 0 \text{ for all } i\in I(\bar{\mathsf{x}})\right\}\,. \tag{10}$$

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.

For concise formulation, 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

$$\Gamma_i = \{\mathsf{v}\in\mathbb{R}^n : \mathsf{D}_i\mathsf{v}\ge\mathsf{o}\}\,,\quad i\in J(\bar{\mathsf{x}})\,. \tag{11}$$

Then $\bigcup_{i\in J(\bar{\mathsf{x}})}\Gamma_i = \Gamma(\bar{\mathsf{x}})$ from (10), and $M_i(\bar{\mathsf{x}}) = M\cap(\Gamma_i+\bar{\mathsf{x}})$ contains all feasible points $\mathsf{x}\in M$ with the following property: $i\in J(\bar{\mathsf{x}})$ is the first inactive constraint to become active when travelling along a ray emanating from $\bar{\mathsf{x}}$ in direction $\mathsf{x}-\bar{\mathsf{x}}$ (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:

$$\mathsf{Q}_i = \mathsf{a}_i\,\nabla f(\bar{\mathsf{x}})^\top + \nabla f(\bar{\mathsf{x}})\,\mathsf{a}_i^\top + \bar{s}_i\mathsf{Q}\,,\quad i\in J(\bar{\mathsf{x}})\,. \tag{12}$$

Theorem 2.2 *For the QP (9), we have that $\bar{\mathsf{x}}$ is a global solution to (9) 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*

$$\widetilde{\mathsf{x}} = \bar{\mathsf{x}} + \frac{\bar{s}_i}{\mathsf{a}_i^\top \mathsf{v}} \mathsf{v} \quad \text{is an improving feasible point,}$$

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 (9) is unbounded.*

Proof. See [1, 6] or also [2, Sect.1.2]. □

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 (11) 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 (12) renders*

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

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* $\Gamma(\bar{\mathsf{x}}) = \bigcup_{i \in J(\bar{\mathsf{x}})} \Gamma_i$*.*

2.3 Illustration: the classical trust region problem

We now illustrate a simple application of the preceding general principles 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:

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

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 $g_1(\mathsf{x}) = \frac{1}{2}(\|\mathsf{x}\|^2 - 1)$ and no equality constraints. If $g_1(\mathsf{x}) = 0$, then the gradient $\nabla g_1(\mathsf{x}) = \mathsf{x}$ 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 $u \geq 0$, 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, as remarked before, does not depend on x. Gay [8] was the first to show that any KKT point x is a global solution to (13) if and only if $\mathsf{H}_{\bar{u}}$ is positive-semidefinite for the unique Lagrange multiplier $\bar{u}$.

First we show that zero multipliers imply global optimality for this problem.

Corollary 2.1 *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 $\|\mathsf{x}\| - 1 = u = 0$). Thus Theorem 1.1(b) implies $\Gamma(\mathsf{x})$-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 $\|\mathsf{x}\| = 1$, then $\Gamma(\bar{\mathsf{x}}) = \left\{\mathsf{v} \in \mathbb{R}^n : \mathsf{v}^\top \bar{\mathsf{x}} \leq 0\right\}$ 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,

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

is a half-space and since $\nabla f(\bar{\mathsf{x}}) = -\bar{u}\bar{\mathsf{x}}$, 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 2.2 *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 $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$ 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 $\mathsf{v}^\top \mathsf{H}_{\bar{u}} \mathsf{v} \ge 0$ 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 $-\lambda_2 \le \bar{u} < -\lambda_1$ and $\mathsf{v}_1^\top \bar{\mathsf{x}} \ne 0$.

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 $\mathsf{v}^\top \mathsf{H}_{\bar{u}} \mathsf{v} \ge 0$ 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 $\mathsf{v}_1^\top \bar{\mathsf{x}} \ne 0$. 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 $\bar{u} \ge -\lambda_1$, 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 $\bar{u} + \lambda_2 \ge 0$ must hold. Hence $-\lambda_2 \le \bar{u} < -\lambda_1$. In particular, $\lambda_1 < \lambda_2$. Since $\mathsf{v}_1^\top \mathsf{H}_{\bar{u}} \mathsf{v}_1 = \lambda_1 + \bar{u} \le \lambda_1 - \lambda_2 < 0$, assertion (a) implies $\mathsf{v}_1^\top \bar{\mathsf{x}} \ne 0$. □

All these results are interesting by themselves, but they also play a crucial role for Martínez' impressive argumentation who proved the following theorem in [11]:

Theorem 2.3 *There is at most one LNG solution to (13).*

Acknowledgements. The author gratefully acknowledges the hospitality, stimulating atmosphere and interesting events offered by the Isaac Newton Institute of 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. I am also indebted to Andreas Fischer, Michael Overton and in particular to Oliver Stein for valuable suggestions and hints.

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] Immanuel M. Bomze and Michael L. Overton. Narrowing the difficulty gap for the Celis-Dennis-Tapia problem. *Math. Program.*, Ser.B, to appear. Also available as: Preprint NI13063–POP, Isaac Newton Institute, Cambridge UK, 2014.

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

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

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

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

[8] David M. Gay. Computing optimal locally constrained steps. *SIAM J. Sci. Stat. Comput.*, 2(2):186–197, 1981.

[9] 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.

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

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

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

[13] Ya-Xiang Yuan. On a subproblem of trust region algorithms for constrained optimization. *Math. Program.*, 47(1, (Ser. A)):53–63, 1990.