Linearized M-stationarity conditions for general optimization problems

Helmut Gfrerer*

Abstract

This paper investigates new first-order optimality conditions for general optimization problems. These optimality conditions are stronger than the commonly used M-stationarity conditions and are in particular useful when the latter cannot be applied because the underlying limiting normal cone cannot be computed effectively. We apply our optimality conditions to a MPEC to demonstrate their practicability.

Key words. M-stationarity conditions; limiting normal cone; regular normal cone; mathematical programs with equilibrium constraints.

Mathematics subject classification. 49J40, 49J52, 90C.

1 Introduction

This paper deals with first-order optimality conditions for general optimization problems of the form

$$\min_z f(z) \quad \text{subject to} \quad P(z) \in D \tag{1}$$

where the mappings $f : \mathbb{R}^d \to \mathbb{R}$ and $P : \mathbb{R}^d \to \mathbb{R}^s$ are assumed to be continuously differentiable and D is a closed subset of $\mathbb{R}^s$. Note that formally more general problems of the form

$$\begin{aligned} \min_z \quad & f(z) \\ \text{subject to} \quad & 0 \in P(z) + Q(z), \end{aligned} \tag{2}$$

where $Q : \mathbb{R}^d \rightrightarrows \mathbb{R}^s$ is a set-valued mapping with closed graph, can be equivalently written in the form (1) as

$$\min f(z) \quad \text{subject to} \quad (z, -P(z)) \in \operatorname{gph} Q. \tag{3}$$

*Institute of Computational Mathematics, Johannes Kepler University (JKU) Linz, A-4040 Linz, Austria, e-mail: helmut.gfrerer@jku.at.

If the objective function in (1) is not continuously differentiable, we can equivalently rewrite the program (1) as

$$\min_{z,\alpha} \alpha \quad \text{subject to} \quad (z, \alpha, P(z)) \in \operatorname{epi} f \times D. \tag{4}$$

Under some constraint qualification, necessary optimality conditions for the problem (1) at a local minimizer $\bar{z}$ are usually of the form

$$0 \in \nabla f(\bar{z}) + \nabla P(\bar{z})^* w^*, \tag{5}$$

where the multiplier w^* belongs to a suitable normal cone to the set D at the point $P(\bar{z})$, which in turn is often related to the notion of a subdifferential. Among the big number of different normal cones/subdifferential constructions considered in the literature, two stand out by the comprehensive calculus available for them: One is given by the *generalized gradient* as introduced by Clarke [4] and the related normal cone, the other one is the *limiting (Mordukhovich) normal cone/subdifferential.* Since the Clarke normal cone is the closure of the convex hull of the limiting normal cone, c.f. [24], the use of the limiting normal cone yields stronger first-order optimality conditions than an approach based on Clarke's normal cone and for this reason we focus in this paper on first-order optimality conditions related to the limiting normal cone, which are usually called M-stationarity conditions. However, despite the available calculus, it is sometimes very difficult or even impossible to compute the limiting normal cone effectively.

As an illustrating example let us consider the following subclass of so-called *mathematical programs with equilibrium constraints* (MPECs), where the equilibrium is described by a generalized equation:

$$\begin{aligned}
\text{(MPEC)} \qquad & \min_{x,y} F(x,y) \\
& \text{s.t. } 0 \in \phi(x,y) + \widehat{N}_\Gamma(y), \\
& \qquad G(x,y) \le 0
\end{aligned} \tag{6}$$

For this problem, the mappings $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $\phi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ and $G : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$ are assumed to be continuously differentiable, $\Gamma := \{y \,|\, g(y) \le 0\}$ is given by a C^2-mapping $g : \mathbb{R}^m \to \mathbb{R}^q$ and $\widehat{N}_\Gamma(y)$ denotes the *regular (Fréchet) normal cone* to Γ at y, cf. Definition 1 below. The program (MPEC) can be equivalently written in the format (1) as

$$\begin{aligned}
\text{(MPEC')} \qquad & \min_{x,y} F(x,y) \\
& \text{s.t. } \hat{P}(x,y) := \begin{pmatrix} (y, -\phi(x,y)) \\ G(x,y) \end{pmatrix} \in \operatorname{gph} \widehat{N}_\Gamma \times \mathbb{R}^p_- =: \hat{D}
\end{aligned} \tag{7}$$

The calculation of the limiting normal cone to $\hat{D}$ at $\hat{P}(\bar{x}, \bar{y})$ involves the one of the limiting normal cone to $\operatorname{gph} \widehat{N}_\Gamma$ at $(\bar{y}, -\phi(\bar{x}, \bar{y}))$. The latter task is well-understood, if for the inequalities $g(y) \le 0$ the *linear independence constraint qualification (LICQ)* is fulfilled at

$\bar{y}$, cf. [18]. The situation, unfortunately, becomes substantially more difficult, provided LICQ is relaxed. Such a situation has been investigated under Mangasarian-Fromovitz constraint qualification (MFCQ) in [16] and, under a certain constraint qualification less restrictive than MFCQ, in [12]. In both cases an additional condition is needed to obtain a point based representation of the limiting normal cone to gph $\widehat{N}_\Gamma$ in terms of first-order and second-order derivatives of g at $\bar{y}$ and in [12] a simple example is given that without this additional condition the limited normal cone cannot be entirely expressed in terms of first-order and second-order derivatives of g.

On the other hand, very recently much progress has been achieved in computing the tangent cone to gph $\widehat{N}_\Gamma$ and to the tangent cone of the feasible region of (6), see [13, 5, 14]. Under very mild assumptions one obtains a full description of the tangent cone to the feasible region of (6) involving only first-order derivatives of ϕ, G and derivatives of g up to second-order at a point $(\bar{x}, \bar{y})$. Thus there must exist also some dual optimality condition in terms of these derivatives showing that the part of the limiting normal cone which is difficult to compute does not play a role in the optimality conditions.

At this point let us mention that it might be not feasible to reformulate the MPEC (6) as a *mathematical program with complementarity constraints (MPCC)*,

$$\begin{aligned} \min_{x,y,\lambda}\ & F(x,y) \\ \text{s.t.}\ & 0 \in \phi(x,y) + \nabla g(y)^* \lambda, \\ & 0 \le \lambda \perp g(y) \ge 0, \\ & G(x,y) \le 0. \end{aligned} \tag{8}$$

Of course, if $(\bar{x}, \bar{y})$ is a local solution of (6) and the system $g(y) \le 0$ fulfills some constraint qualification at $\bar{y}$ ensuring $\widehat{N}_\Gamma(\bar{y}) = \{\nabla g(\bar{y})^* \lambda \,|\, 0 \le \lambda \perp g(\bar{y})\}$, then it is easy to show that for every multiplier $\bar{\lambda} \ge 0$ fulfilling $0 \in \phi(\bar{x}, \bar{y}) + \nabla g(\bar{y})^* \bar{\lambda}$, $\bar{\lambda}^T g(\bar{y}) = 0$ the triple $(\bar{x}, \bar{y}, \bar{\lambda})$ is a local solution of (8). However, if LICQ fails to hold for the system $g(y) \le 0$ at $\bar{y}$, then it can happen that some constraint qualification is fulfilled for the MPEC (6), but all of the MPCC-tailored constraint qualifications known from the literature are violated for (8). Thus we cannot apply the known first-order optimality conditions for the program (8) in order to obtain optimality conditions for the program (6). This was first observed in [1] and further developed in [14]. In the latter paper an example is given where this phenomena occurs for convex quadratic functions g_i, $i = 1, \ldots, q$ and linear mappings ϕ and G.

To overcome the difficulties arising when computing the limiting normal cone, we remember that the basic task in formulating first-order optimality conditions is the computation of the regular normal cone to the feasible set of (1). However, for the regular normal cone only very restricted calculus is available and this is the reason why the limiting normal cone is used instead of the regular one. Having in mind that the basic goal is the computation of the regular normal cone to the feasible set, it is not difficult to see that in order to obtain a more accurate approximation we can use the limiting normal cone to the tangent cone of the feasible set. Performing a more accurate analysis we observe that this process can be repeated and we obtain as a final result that the multiplier w^* in (5) is a regular normal to a series of tangent cones to tangent cones to the set D. Since the new optimality conditions are derived by a repeated linearization procedure, we call the resulting optimality conditions *linearized M-stationarity conditions*.

The organization of the paper is as follows. In Section 2 we recall some basics from variational analysis. The stationarity concepts of B-,S- and M-stationarity and its relations with necessary optimality conditions are considered in Section 3.

Section 4 contains the main results on linearized M-stationarity conditions for the problem (1). The analysis is done under a very weak constraint qualification: We only require the *generalized Guignard constraint qualification (GGCQ)* and the *metric subregularity constraint qualification (MSCQ)* for the linearized problem. In particular, both conditions are fulfilled if MSCQ holds for the problem (1).

We apply these results to the MPEC (6) in Section 5 and derive the linearized M-stationarity conditions under a certain condition on the lower level system $q_i(y) \le 0$, $i = 1, \dots, p$, which is weaker than the *constant rank constraint qualification* (CRCQ). This also works when we are not able to compute the limiting normal cone to gph $\widehat{N}_\Gamma$ as in [12].

In the concluding Section 6 we briefly summarize the obtained results and outline some topics for our future research.

Throughout the paper we use standard notation of variational analysis and generalized differentiation. For an element $z \in \mathbb{R}^d$ we denote by $[z]$ the subspace $\{\alpha z \mid \alpha \in \mathbb{R}\}$ generated by z. Some more special symbols are introduced when appearing first in the text.

2 Preliminaries from variational analysis

All the sets under consideration are supposed to be locally closed around the points in question without further mentioning. We recall first the standard constructions of variational analysis used in what follows.

Definition 1. *Given a set $\Omega \subseteq \mathbb{R}^d$ and a point $\bar{z} \in \Omega$, the (Bouligand-Severi)* tangent/contingent cone *to Ω at $\bar{z}$ is a closed cone defined by*

$$T_\Omega(\bar{z}) := \left\{ w \in \mathbb{R}^d \,\middle|\, \exists t_k \downarrow 0,\ w_k \to w \ \ \textit{with}\ \ \bar{z} + t_k w_k \in \Omega\ \forall\ k \right\}.$$

The (Fréchet) regular normal cone *and the (Mordukhovich)* limiting/basic normal cone *to Ω at $\bar{z}$ are defined by*

$$\begin{aligned} &\widehat{N}_\Omega(\bar{z}) := (T_\Omega(\bar{z}))^* \\ \textit{and} \qquad &N_\Omega(\bar{z}) := \left\{ z^* \mid \exists z_k \stackrel{\Omega}{\to} \bar{z} \ \textit{and}\ z_k^* \to z^* \ \textit{such that}\ z_k^* \in \widehat{N}_\Omega(z_k)\ \forall k \right\} \end{aligned}$$

respectively.
Further, if $\bar{z} \notin \Omega$ we define

$$T_\Omega(\bar{z}) := \widehat{N}_\Omega(\bar{z}) := N_\Omega(\bar{z}) := \emptyset.$$

When the set Ω is convex, the tangent/contingent cone and the regular/limiting normal cone reduce to the classical tangent cone and normal cone of convex analysis respectively. The regular normal cone $\widehat{N}_\Omega(\bar{z})$ is always convex whereas the limiting normal cone can be non-convex if Ω is not convex.

Lemma 1. *Let $\Omega \subseteq \mathbb{R}^d$ be closed and $\bar{z} \in \Omega$. Then*

$$N_\Omega(\bar{z}) \supseteq N_{T_\Omega(\bar{z})}(0) = \widehat{N}_\Omega(\bar{z}) \cup \bigcup_{0 \neq w \in \mathbb{R}^d} N_{T_\Omega(\bar{z})}(w). \tag{9}$$

Proof. The inclusion $N_\Omega(\bar{z}) \supseteq N_{T_\Omega(\bar{z})}(0)$ in (9) was shown in [24, Proposition 6.27]. It also follows from [24, Proposition 6.27] together with $N_{T_\Omega(\bar{z})}(0) \supseteq \widehat{N}_{T_\Omega(\bar{z})}(0) = \widehat{N}_\Omega(\bar{z})$ that $N_{T_\Omega(\bar{z})}(0) \supseteq \widehat{N}_\Omega(\bar{z}) \cup \bigcup_{0 \neq w \in \mathbb{R}^d} N_{T_\Omega(\bar{z})}(w)$. In order to show the reverse inclusion, consider $w^* \in N_{T_\Omega(\bar{z})}(0)$ together with sequences $w_k \to 0$, ${w_k}^* \to w^*$ with $w_k^* \in \widehat{N}_{T_\Omega(\bar{z})}(w_k)$ $\forall k$. If $w_k = 0$ holds for infinitely many k, then $w^* \in \widehat{N}_{T_\Omega(\bar{z})}(0) = \widehat{N}_\Omega(\bar{z})$ follows because $\widehat{N}_\Omega(\bar{z})$ is closed. On the other hand, if $w_k \neq 0$ holds for all but finitely many k by passing to a subsequence we can assume that $w_k/\|w_k\|$ converges to some w, and because of $\widehat{N}_{T_\Omega(\bar{z})}(w_k) = \widehat{N}_{T_\Omega(\bar{z})}(w_k/\|w_k\|)$ we conclude $w^* \in N_{T_\Omega(\bar{z})}(w)$. Hence (9) is established and this finishes the proof. $\square$

Usually, the computation of the limiting normal cone to a nonconvex set Ω is a difficult task. A special case when the limiting normal cone has a comparatively simple description is given by polyhedral sets.

Definition 2. *Let $\Omega \subseteq \mathbb{R}^d$.*

1. *We say that Ω is* convex polyhedral*, if the set can be written as the intersection of finitely many halfspaces, i.e. there are elements $(a_i, \alpha_i) \in \mathbb{R}^d \times \mathbb{R}$, $i = 1, \dots, p$ such that $\Omega = \{z \,|\, \langle a_i, z\rangle \leq \alpha_i,\ i = 1, \dots, p\}$.*

2. *We say that Ω is* polyhedral*, if it is the union of finitely many convex polyhedral sets.*

3. *Given a point $\bar{z} \in \Omega$, we say that Ω is* locally polyhedral near $\bar{z}$ *if there is a neighborhood W of $\bar{z}$ and a polyhedral set C such that $\Omega \cap W = C \cap W$.*

Lemma 2. *Let $\Omega \subseteq \mathbb{R}^d$ be locally polyhedral near some point $\bar{z} \in \Omega$. Then*

$$N_\Omega(\bar{z}) = \bigcup_{w \in T_\Omega(\bar{z})} \widehat{N}_{T_\Omega(\bar{z})}(w). \tag{10}$$

Proof. Follows from [10, Lemma 2.2]. $\square$

In this paper the notion of *metric subregularity* will play an important role.

Definition 3. *Let $M : \mathbb{R}^d \rightrightarrows \mathbb{R}^s$ be a set-valued mapping and let $(\bar{z}, \bar{w}) \in \operatorname{gph} M$. We say that M is* metrically subregular *at $(\bar{z}, \bar{w})$ if there exist a neighborhood W of $\bar{z}$ and a positive number $\kappa > 0$ such that*

$$\mathrm{d}(z, M^{-1}(\bar{w})) \le \kappa \mathrm{d}(\bar{w}, M(z)) \quad \forall z \in W. \tag{11}$$

It is well-known that metric subregularity of M at $(\bar{z}, \bar{w})$ is equivalent with the property of *calmness* of the inverse mapping M^{-1} at $(\bar{w}, \bar{z})$, cf. [7]. Further, metric subregularity of M at $(\bar{z}, \bar{w})$ is equivalent with metric subregularity of the mapping $z \to (z, \bar{w}) - \operatorname{gph} M$ at $(\bar{z}, (0, 0))$, cf. [14, Proposition 3].

Lemma 3. *Let $M : \mathbb{R}^d \rightrightarrows \mathbb{R}^s$ be a set-valued mapping, let $(\bar{z}, \bar{w}) \in \operatorname{gph} M$ and assume that* $\operatorname{gph} M$ *is a closed cone. If M is metrically subregular at $(0, 0)$ then there is some $\kappa > 0$ such that*

$$\mathrm{d}(z, M^{-1}(0)) \le \kappa \mathrm{d}(0, M(z)) \ \forall z \in \mathbb{R}^d.$$

In particular, M is metrically subregular at every point $(\bar{z}, 0) \in \operatorname{gph} M$.

Proof. According to the definition of metric subregularity, consider a neighborhood W of 0 and a real $\kappa > 0$ such that $\mathrm{d}(z, M^{-1}(0)) \le \kappa \mathrm{d}(0, M(z))$ for all $z \in W$. Now consider $z \in \mathbb{R}^d$. Then we can find some $\lambda > 0$ such that $\lambda z \in W$ and thus $\mathrm{d}(\lambda z, M^{-1}(0)) \le \kappa \mathrm{d}(0, M(\lambda z))$. Since $\operatorname{gph} M$ is a cone it follows that $M^{-1}(0)$ is a cone and $M(\lambda z) = \lambda M(z)$. Hence $\lambda \mathrm{d}(z, M^{-1}(0)) = \mathrm{d}(\lambda z, M^{-1}(0)) \le \kappa \mathrm{d}(0, M(\lambda z)) = \lambda \kappa \mathrm{d}(0, M(z))$ and $\mathrm{d}(z, M^{-1}(0)) \le \kappa \mathrm{d}(0, M(z))$ follows. $\square$

The following lemma is a special variant of [9, Proposition 2.1].

Lemma 4. *Let $P : \mathbb{R}^d \to \mathbb{R}^s$ be contiunously differentiable, let $D \subseteq \mathbb{R}^s$ be closed and assume that the mapping $z \rightrightarrows P(z) - D$ is metrically subregular at $(\bar{z}, 0)$. Then the mapping $u \rightrightarrows \nabla P(\bar{z})u - T_D(P(\bar{z}))$ is metrically subregular at $(0, 0)$.*

Given a cone $C \subseteq \mathbb{R}^d$, we denote by $\mathcal{L}(C)$ the largest subspace $L \subseteq \mathbb{R}^d$ such that

$$C + L \subseteq C.$$

Note that $\mathcal{L}(C)$ is well defined because for two subspaces L_1, L_2 fulfilling $C + L_i \subseteq C$, $i = 1, 2$ we have

$$C + L_1 + L_2 = (C + L_1) + L_2 \subseteq C + L_2 \subseteq C \tag{12}$$

and we are working in finite dimensional spaces. Note that for every subspace L we have $C + L \supseteq C$ and thus $C + \mathcal{L}(C) = C$. If C is a convex cone, then $\mathcal{L}(C) = C \cap (-C)$ is the so-called *lineality space* of C, the largest subspace contained in C.

Lemma 5. *Let $C \subseteq \mathbb{R}^d$ be a closed cone and let $\bar{z} \in C$. Then*

$$\mathcal{L}(C) + [\bar{z}] \subseteq \mathcal{L}(T_C(\bar{z})).$$

Proof. We show that both $T_C(\bar{z}) + \mathcal{L}(C) \subseteq T_C(\bar{z})$ and $T_C(\bar{z}) + [\bar{z}] \subseteq T_C(\bar{z})$. Then the statement follows from (12). Consider a tangent $w \in T_C(\bar{z})$ together with sequences $t_k \downarrow 0$ and $w_k \to w$ with $\bar{z} + t_k w_k \in C$ for all k. For fixed $l \in \mathcal{L}(C)$ and for every k we have $t_k l \in \mathcal{L}(C)$ and thus $\bar{z} + t_k w_k + t_k l = \bar{z} + t_k(w_k + l) \in C$. Hence $w + l \in T_C(\bar{z})$ and $T_C(\bar{z}) + \mathcal{L}(C) \subseteq T_C(\bar{z})$ follows. Next, let $\gamma \in \mathbb{R}$. By passing to a subsequence we can assume $1 + t_k\gamma > 0$ and thus

$$(1 + t_k\gamma)(\bar{z} + t_k w_k) = \bar{z} + t_k(1 + t_k\gamma)\left(w_k + \frac{\gamma}{1 + t_k\gamma}\bar{z}\right) \in C \ \forall k.$$

Since $t_k(1 + t_k\gamma) \downarrow 0$ and $w_k + \frac{\gamma}{1+t_k\gamma}\bar{z} \to w + \gamma\bar{z}$, we conclude $w + \gamma\bar{z} \in T_C(\bar{z})$ and the second claimed inclusion $T_C(\bar{z}) + [\bar{z}] \subseteq T_C(\bar{z})$ follows. This finishes the proof. □

At the end of this section we recall the definition of the critical cone to a set.

Definition 4. *Given a set Ω and an element $\bar{z} \in \Omega$ together with a regular normal $\bar{z}^* \in \widehat{N}_\Omega(\bar{z})$ we define the* critical cone *to Ω at $(\bar{z}, \bar{z}^*)$ as*

$$\mathcal{K}_\Omega(\bar{z}, \bar{z}^*) := T_\Omega(\bar{z}) \cap [\bar{z}^*]^\perp.$$

3 Stationarity concepts

In this section we recall some basic fact about stationarity concepts for the general problem (1).

We denote by Ω the feasible region of the problems (1), i.e.

$$\Omega \;:=\; \{z \in \mathbb{R}^d \,|\, P(z) \in D\} \tag{13}$$

Further, given $\bar{z} \in \Omega$ we denote by

$$T^{\text{lin}}_{P,D}(\bar{z}) := \{u \in \mathbb{R}^d \,|\, \nabla P(\bar{z})u \in T_D(P(\bar{z}))\}$$

the *linearized tangent cone to Ω at $\bar{z}$.* Recall that there always holds

$$T_\Omega(\bar{z}) \subseteq T^{\text{lin}}_{P,D}(\bar{z}). \tag{14}$$

We use the notation $T^{\text{lin}}_{P,D}(\bar{z})$ to indicate that the linearized tangent cone depends on P and D, i.e., if we have two equivalent representations

$$\Omega = \{z \,|\, P_1(z) \in D_1\} = \{z \,|\, P_2(z) \in D_2\} \tag{15}$$

with continuously differentiable mappings $P_i : \mathbb{R}^d \to \mathbb{R}^{s_i}$ and closed sets $D_i \subseteq R^{s_i}$, $i = 1, 2$, then we can have $T^{\text{lin}}_{P_1,D_1}(\bar{z}) \neq T^{\text{lin}}_{P_2,D_2}(\bar{z})$.

Definition 5. *Let $\bar{z} \in \Omega$. We say that $\bar{z}$ is*

1. B-stationary (Bouligand stationary) *for the problem* (1)*, if*

$$0 \in \nabla f(\bar{z}) + \widehat{N}_\Omega(\bar{z}),$$

2. S-stationary (strong stationary) *for the problem* (1)*, if*

$$0 \in \nabla f(\bar{z}) + \nabla P(\bar{z})^* \widehat{N}_D(P(\bar{z})),$$

3. M-stationary *for the problem* (1)*, if*

$$0 \in \nabla f(\bar{z}) + \nabla P(\bar{z})^* N_D(P(\bar{z})).$$

Note that S- and M-stationarity depend on P and D used for describing of Ω whereas B-stationarity is independent of the representation of Ω.

B-stationarity can be equivalently expressed as

$$\langle \nabla f(\bar{z}), w \rangle \geq 0 \ \forall w \in T_\Omega(\bar{z}).$$

By saying that a *feasible descent direction* for the program (1) at $\bar{z}$ is a direction $w \in T_\Omega(\bar{z})$ with $\langle \nabla f(\bar{z}), w \rangle < 0$, we see that B-stationarity conveys the fact that no feasible descent direction exists. It is well known that every local minimizer is also B-stationary, cf. [24, Theorem 6.12]. Conversely, if $\bar{z}$ is B-stationary for the program (1), then by [24, Theorem 6.11] there exists a smooth mapping $\tilde{f} : \mathbb{R}^d \to \mathbb{R}$ such that $\tilde{f}(\bar{z}) = f(\bar{z})$, $\nabla \tilde{f}(\bar{z}) = \nabla f(\bar{z})$ and $\bar{z}$ is a global minimizer of the program

$$\min_z \tilde{f}(z) \quad \text{subject to} \quad P(z) \in D.$$

Thus, if the available first-order information at the point $\bar{z}$ is provided solely by $T_\Omega(\bar{z})$ and $\nabla f(\bar{z})$, then B-stationarity constitutes the best possible first-order optimality condition and thus characterizing B-stationarity is the primary goal.

However, the computation of the regular normal cone $\widehat{N}_\Omega(\bar{z})$ appearing in the definition of B-stationarity can be a very difficult task for general sets D and therefore, besides other stationary concepts, the notions of S- and M-stationarity have been introduced. S-stationarity was first considered in the monograph by Luo, Pang and Ralph [17] whereas M-stationarity conditions appeared first in the papers by Outrata [19] and Ye [25], respectively. The monikers M-stationarity and S-stationarity were coined in [20, 21] for MPCC and then carried over in [8] to the general problem (1).

By applying [24, Theorem 6.14] we readily obtain the inclusion

$$\widehat{N}_\Omega(\bar{z}) \supseteq \nabla P(\bar{z})^* \widehat{N}_D(P(\bar{z})). \tag{16}$$

Hence we deduce from the definition that S-stationarity of $\bar{z}$ implies B-stationarity. However, the reverse implication is only valid under comparatively strong assumptions. We state here the following result due to Gfrerer and Outrata [13, Theorem 4].

Theorem 1. *Assume that $\bar{z}$ is feasible for the problem* (1)*, assume that the mapping $z \rightrightarrows P(z) - D$ is metrically subregular at $(\bar{z}, 0)$ and assume that*

$$\nabla P(\bar{z})\mathbb{R}^d + \mathcal{L}(T_D(P(\bar{z}))) = \mathbb{R}^s.$$

Then (16) *holds with equality. In particular, if $\bar{z}$ is B-stationary then it is S-stationary as well.*

It is well known that B-stationarity implies M-stationarity under mild constraint qualification conditions.

Definition 6. *Let $P(\bar{z}) \in D$.*

1. *(cf. [8]) We say that the* generalized Abadie constraint qualification *(GACQ) holds at $\bar{z}$ if*
$$T_\Omega(\bar{z}) = T^{\text{lin}}_{P,D}(\bar{z}). \tag{17}$$

2. *(cf. [8]) We say that the* generalized Guignard constraint qualification *(GGCQ) holds at $\bar{z}$ if*
$$\widehat{N}_\Omega(\bar{z}) = \left(T^{\text{lin}}_{P,D}(\bar{z})\right)^*. \tag{18}$$

3. *(cf. [11]) We say that the* metric subregularity constraint qualification (MSCQ) *holds at $\bar{z}$ if the set-valued map $M(z) := P(z) - D$ is metrically subregular at $(\bar{z}, 0)$.*

We always have

$$\text{MSCQ} \implies \text{GACQ} \implies \text{GGCQ}.$$

Indeed, the first implication follows from [15, Proposition 1] whereas the second implication obviously holds true. Note that all these constraint qualifications depend on the representation of Ω by P and D. GGCQ seems to be indispensable for verifying B-stationarity solely with first-order derivatives of the problem functions.

We state here the following result from the recent paper by Benko and Gfrerer [2, Proposition 3].

Theorem 2. *Assume that $\bar{z}$ is feasible for the problem* (1) *and assume that GGCQ is fulfilled, while the mapping $u \rightrightarrows \nabla P(\bar{z})u - T_D(P(\bar{z}))$ is metrically subregular at* $(0, 0)$*. Then*

$$\widehat{N}_\Omega(\bar{z}) \subseteq \nabla P(\bar{z})^* N_{T_D(P(\bar{z}))}(0) \subseteq \nabla P(\bar{z})^* N_D(P(\bar{z})). \tag{19}$$

Remark 1. *Note that the assumptions of Theorem 2 are fulfilled if MSCQ holds at $\bar{z}$. Indeed, MSCQ implies GGCQ and metric subregularity of $u \rightrightarrows \nabla P(\bar{z})u - T_D(P(\bar{z}))$ at* $(0, 0)$ *follows from Lemma 4.*

If $\bar{z}$ is B-stationary and the assumptions of Theorem 2 are fulfilled, it follows from the second inclusion in (19) and the definition that $\bar{z}$ is M-stationary. Other constraint qualifications ensuring M-stationarity can be found in [26]. However, from the first inclusion in (19) we also derive the necessary optimality condition

$$0 \in \nabla f(\bar{z}) + \nabla P(\bar{z})^* N_{T_D(P(\bar{z}))}(0) \tag{20}$$

and this is stronger than M-stationarity because we always have

$$N_{T_D(P(\bar z))}(0) \subseteq N_D(P(\bar z))$$

by [24, Proposition 6.27].

4 Linearized M-stationarity conditions

One of the basic statements of this section is provided by the following proposition, which can be considered as a refinement of the necessary condition (20).

Proposition 1. *Let $\bar z$ be B-stationary for the optimization problem* (1) *and assume that GGCQ is fulfilled, while the mapping $u \rightrightarrows \nabla P(\bar z)u - T_D(P(\bar z))$ is metrically subregular at* $(0,0)$. *Then one of the following two conditions is fulfilled:*

1. *There is $w \in T_D(P(\bar z))$ and a multiplier $w^* \in \widehat{N}_{T_D(P(\bar z))}(w)$ such that*

$$\nabla f(\bar z) + \nabla P(\bar z)^* w^* = 0. \tag{21}$$

2. *There is $\bar u \in T^{\mathrm{lin}}_{P,D}(\bar z)$ such that*

$$\nabla P(\bar z)\bar u \notin \mathcal{L}(T_D(P(\bar z))), \tag{22}$$
$$\langle \nabla f(\bar z), \bar u\rangle = 0, \tag{23}$$
$$0 \in \nabla f(\bar z) + \widehat{N}_{T^{\mathrm{lin}}_{P,D}(\bar z)}(\bar u) \tag{24}$$

and $T_D(P(\bar z))$ is not locally polyhedral near $\nabla P(\bar z)\bar u$.

Before proving this theorem we discuss some of its issues. We will call a direction $u \in T^{\mathrm{lin}}_{P,D}(\bar z)$ satisfying (23) a *critical direction* for the problem (1). Now assume that the first statement of Proposition 1 fails to hold and thus there exist $\bar u$ fulfilling the second statement. Let us rename $\bar u$ by u_1. From (22) it follows that $\nabla P(\bar z)u_1 \neq 0$ and thus $u_1 \neq 0$ as well. Further, since u_1 is a critical direction and $\bar z$ is assumed to be B-stationary for the problem (1), it follows that u_1 is a global solution of the program

$$\min \langle \nabla f(\bar z), u\rangle \quad \text{subject to} \quad \nabla P(\bar z)u \in T_D(P(\bar z)) \tag{25}$$

and (24) is the corresponding B-stationarity condition. This is not really surprising, but the important point is that we can apply Proposition 1 once more to the problem (25) at u_1. Indeed, since the mapping $u \rightrightarrows \nabla P(\bar z)u - T_D(P(\bar z))$ is assumed to be metrically subregular at $(0,0)$ and its graph is a closed cone, by Lemma 3 it is metrically subregular at $(u_1, 0)$ as well. By taking into account Remark 1 we see that GGCQ holds for the system $\nabla P(\bar z)u \in T_D(P(\bar z))$ at u_1 and the linearized mapping $u \rightrightarrows \nabla P(\bar z)u - T_{T_D(P(\bar z))}(\nabla P(\bar z)u_1)$ is metrically subregular at $(0,0)$. Thus we can apply Proposition 1 to obtain either the existence of some direction $w \in T_{T_D(P(\bar z))}(\nabla P(\bar z)u_1)$ and some multiplier $w^* \in \widehat{N}_{T_{T_D(P(\bar z))}(\nabla P(\bar z)u_1)}(w)$

such that (21) holds or the existence of some direction $u_2 \in T^{\text{lin},1}_{P,D}(\bar{z}; u_1) := \{u \,|\, \nabla P(\bar{z})u \in T_{T_D(P(\bar{z}))}(\nabla P(\bar{z})u_1)\}$ such that

$$\begin{aligned}
&\nabla P(\bar{z})u_2 \notin \mathcal{L}(T_{T_D(P(\bar{z}))}(\nabla P(\bar{z})u_1)),\\
&\langle \nabla f(\bar{z}), u_2\rangle = 0,\\
&0 \in \nabla f(\bar{z}) + \widehat{N}_{T^{\text{lin}}_{\nabla P(\bar{z}),T_D(P(\bar{z}))}(u_1)}(u_2)
\end{aligned}$$

and $T_{T_D(P(\bar{z}))}(\nabla P(\bar{z})u_1)$ is not locally polyhedral near $\nabla P(\bar{z})u_2$. Again, if the first case does not emerge we can repeat the procedure. Let us recursively define for $\bar{y} \in D$ and directions $v_1, v_2, \ldots$ the following k-th order tangent cones to D by

$$T^0_D(\bar{y}) := T_D(\bar{y}),\ T^k_D(\bar{y}; v_1, \ldots, v_k) := T_{T^{k-1}_D(\bar{y};v_1,\ldots,v_{k-1})}(v_k),\ k \geq 1.$$

Note that by the definition of the tangent cone we have $T^k_D(\bar{y}; v_1, \ldots, v_k) = \emptyset$ if $v_k \notin T^{k-1}_D(\bar{y}; v_1, \ldots, v_{k-1})$. Then we can also define the following k-th order linearized tangent cones to Ω by

$$\begin{aligned}
&T^{\text{lin},0}_{P,D}(\bar{z}) = T^{\text{lin}}_{P,D}(\bar{z}),\\
&T^{\text{lin},k}_{P,D}(\bar{z}; u_1, \ldots, u_k) := \{u \,|\, \nabla P(\bar{z})u \in T^k_D(P(\bar{z}); \nabla P(\bar{z})u_1, \ldots, \nabla P(\bar{z})u_k)\},\ k \geq 1.
\end{aligned}$$

When we apply Proposition 1 the k-th time we find either a direction

$$w \in T^{k-1}_D(P(\bar{z}); \nabla P(\bar{z})u_1, \ldots, \nabla P(\bar{z})u_{k-1})$$

together with a multiplier

$$w^* \in \widehat{N}_{T^{k-1}_D(P(\bar{z});\nabla P(\bar{z})u_1,\ldots,\nabla P(\bar{z})u_{k-1})}(w)$$

such that $\nabla f(\bar{z}) + \nabla P(\bar{z})^* w^* = 0$ or a direction $u^k \in T^{\text{lin},k-1}_{P,D}(\bar{z}; u_1, \ldots, u_{k-1})$ such that

$$\nabla P(\bar{z})u_k \notin \mathcal{L}(T^{k-1}_D(P(\bar{z}); \nabla P(\bar{z})u_1, \ldots, \nabla P(\bar{z})u_{k-1})), \tag{26}$$

$$\langle \nabla f(\bar{z}), u_k\rangle = 0, \tag{27}$$

$$0 \in \nabla f(\bar{z}) + \widehat{N}_{T^{\text{lin},k-1}_{P,D}(\bar{z};u_1,\ldots,u_{k-1})}(u_k) \tag{28}$$

and $T^{k-1}_D(P(\bar{z}); \nabla P(\bar{z})u_1, \ldots, \nabla P(\bar{z})u_{k-1})$ is not locally polyhedral near $\nabla P(\bar{z})u_k$. Next observe that we cannot infinitely often apply Proposition 1. By Lemma 5 we have

$$\begin{aligned}
&\mathcal{L}(T^k_D(P(\bar{z}); \nabla P(\bar{z})u_1, \ldots, \nabla P(\bar{z})u_k))\\
&\supseteq \mathcal{L}(T^{k-1}_D(P(\bar{z}); \nabla P(\bar{z})u_1, \ldots, \nabla P(\bar{z})u_{k-1})) + [\nabla P(\bar{z})u_k]
\end{aligned}$$

and together with (26) we obtain

$$\begin{aligned}
&\dim \mathcal{L}(T^k_D(P(\bar{z}); \nabla P(\bar{z})u_1, \ldots, \nabla P(\bar{z})u_k))\\
&\geq \dim \mathcal{L}(T^{k-1}_D(P(\bar{z}); \nabla P(\bar{z})u_1, \ldots, \nabla P(\bar{z})u_{k-1})) + 1.
\end{aligned}$$

Since we work in finite dimensions the finiteness of k follows. Summing up we have shown the following theorem.

Theorem 3. *Let $\bar{z}$ be B-stationary for the optimization problem* (1) *and assume that GGCQ is fulfilled, while the mapping $u \rightrightarrows \nabla P(\bar{z})u - T_D(P(\bar{z}))$ is metrically subregular at $(0,0)$. Then there exists a natural number $k \geq 0$, directions $u_1, \ldots, u_k$ and $w \in T_D^k(P(\bar{z}); \nabla P(\bar{z})u_1, \ldots, \nabla P(\bar{z})u_k)$ and a multiplier $w^* \in \widehat{N}_{T_D^k(P(\bar{z}); \nabla P(\bar{z})u_1, \ldots, \nabla P(\bar{z})u_k)}(w)$ such that*

$$\nabla f(\bar{z}) + \nabla P(\bar{z})^* w^* = 0.$$

Moreover, for every $l = 1, \ldots, k$ we have

$$u_l \in T_{P,D}^{\mathrm{lin},l-1}(\bar{z}; u_1, \ldots, u_{l-1}) \tag{29}$$

$$\nabla P(\bar{z})u_l \notin \mathcal{L}(T_D^{l-1}(P(\bar{z}); \nabla P(\bar{z})u_1, \ldots, \nabla P(\bar{z})u_{l-1})), \tag{30}$$

$$\langle \nabla f(\bar{z}), u_l \rangle = 0 \tag{31}$$

and $T_D^{l-1}(P(\bar{z}); \nabla P(\bar{z})u_1, \ldots, \nabla P(\bar{z})u_{l-1})$ is not locally polyhedral near $\nabla P(\bar{z})u_l$.

It is easy to see that Theorem 3 considerably strengthen the necessary optimality condition (20), which in turn is stronger than the usual M-stationary condition. As candidates for the multipliers w^* fulfilling the first-order optimality condition (5) we consider multipliers fulfilling

$$w^* \in \widehat{N}_{T_D^k(P(\bar{z}), \nabla P(\bar{z})u_1, \ldots, \nabla P(\bar{z})u_k)}(w) \tag{32}$$

for some $w \in T_D^k(P(\bar{z}), \nabla P(\bar{z})u_1, \ldots, \nabla P(\bar{z})u_k)$, where the directions u_l, $l = 1, \ldots, k$ fulfill the conditions of Theorem 3. By applying the following lemma we immediately obtain that the set on the right hand side of the inclusion (32) is contained in $N_{T_D(P(\bar{z}))}(0) \subseteq N_D(P(\bar{z}))$.

Lemma 6. *Let $\bar{y} \in D$. Then for every collection of directions $v_1, \ldots, v_l \in \mathbb{R}^s$ we have*

$$\widehat{N}_{T_D^{l-1}(\bar{y}; v_1, \ldots, v_{l-1})}(v_l) \subseteq N_{T_D^{l-1}(\bar{y}; v_1, \ldots, v_{l-1})}(v_l) \subseteq N_{T_D(\bar{y})}(0) \subseteq N_D(\bar{y}).$$

Proof. We will show the lemma by induction with respect to the number of directions l. Indeed, for $l = 1$ the claimed inclusions hold true because for all v_1 we have $\widehat{N}_{T_D^0(\bar{y})}(v_1) \subseteq N_{T_D^0(\bar{y})}(v_1) \subseteq N_{T_D(\bar{y})}(0) \subseteq N_D(\bar{y})$ by the definitions of the regular/limiting normal cone and (9). Now assume that the claim holds true for some number $l \geq 1$ and consider arbitrary directions $v_1, \ldots, v_{l+1}$. Then by the definitions of the regular/limiting normal cone, (9) and the induction hypothesis we obtain

$$\begin{aligned}
\widehat{N}_{T_D^l(\bar{y}; v_1, \ldots, v_l)}(v_{l+1}) &\subseteq N_{T_D^l(\bar{y}; v_1, \ldots, v_l)}(v_{l+1}) = N_{T_{T_D^{l-1}(\bar{y}; v_1, \ldots, v_{l-1})}(v_l)}(v_{l+1}) \\
&\subseteq N_{T_{T_D^{l-1}(\bar{y}; v_1, \ldots, v_{l-1})}(v_l)}(0) \subseteq N_{T_D^{l-1}(\bar{y}; v_1, \ldots, v_{l-1})}(v_l) \\
&\subseteq N_{T_D(\bar{y})}(0) \subseteq N_D(\bar{y})
\end{aligned}$$

and the lemma is proved. $\square$

We do not know so much about the order k appearing in Theorem 3. By using (30) and Lemma 5, a rough upper estimate for k is given by $\dim(\nabla P(\bar{z})\mathbb{R}^d) - \dim(\mathcal{L}(T_D(P(\bar{z})) \cap$

$\nabla P(\bar{z})\mathbb{R}^d$). However, in many examples we found that this bound is too pessimistic and the necessary optimality conditions of Theorem 3 hold with small k, say $k = 0, 1$ or 2. More research has to be done to investigate this circumstance.

Recall that a local minimizer $\bar{z}$ for (1) is called a *sharp minimum* if there is a constant $\alpha > 0$ such that

$$f(z) \geq f(\bar{z}) + \alpha\|z - \bar{z}\|$$

holds for all feasible z close to $\bar{z}$.

Lemma 7. *Assume that at $\bar{z}$ GGCQ is fulfilled. Then $\bar{z}$ is a sharp minimum if and only if there is some $\alpha' > 0$ such that*

$$\langle \nabla f(\bar{z}), u\rangle \geq \alpha'\|u\| \;\forall u \in T^{\rm lin}_{P,D}(\bar{z}). \tag{33}$$

Proof. In order to show the sufficiency of (33) for $\bar{z}$ being a sharp minimum, assume on the contrary that there is a sequence z_k of feasible points converging to $\bar{z}$ satisfying

$$\liminf_{k\to\infty} \frac{f(z_k) - f(\bar{z})}{\|z_k - \bar{z}\|} = \liminf_{k\to\infty} \langle \nabla f(\bar{z}), \frac{z_k - \bar{z}}{\|z_k - \bar{z}\|}\rangle \leq 0$$

By passing to a subsequence we can assume that $\frac{z_k-\bar{z}}{\|z_k-\bar{z}\|}$ converges to some u. Then $\langle \nabla f(\bar{z}), u\rangle \leq 0$ and $u \in T_\Omega(\bar{z}) \subset T^{\rm lin}_{P,D}(\bar{z})$ contradicting (33). To prove necessity of (33), assume that $\bar{z}$ is a sharp minimum and consider a tangent $u \in T_\Omega(\bar{z})$ together with sequences $t_k \downarrow 0$ and $u_k \to u$ satisfying $P(\bar{z} + t_k u_k) \in D$. Then

$$f(\bar{z} + t_k u_k) - f(\bar{z}) = t_k\langle \nabla f(\bar{z}), u_k\rangle + o(t_k\|u_k\|) \geq \alpha t_k\|u_k\|$$

and by dividing by t_k and passing to the limit we obtain $\langle \nabla f(\bar{z}), u\rangle \geq \alpha\|u\|$. Next consider $u \in \operatorname{conv} T_\Omega(\bar{z})$ together with elements $u_1, \ldots u_K$ and positive scalars $\gamma_1, \ldots, \gamma_K$, $\sum_{i=1}^K \gamma_i = 1$ such that $u = \sum_{i=1}^K \gamma_i u_i$. Then

$$\langle \nabla f(\bar{z}), u\rangle = \sum_{i=1}^K \gamma_i \langle \nabla f(\bar{z}), u_i\rangle \geq \alpha \sum_{i=1}^K \gamma_i\|u_i\| \geq \alpha\|\sum_{i=1}^K \gamma_i u_i\| = \alpha\|u\|$$

and we easily conclude

$$\langle \nabla f(\bar{z}), u\rangle \geq \alpha\|u\| \;\forall u \in \operatorname{cl}\operatorname{conv} T_\Omega(\bar{z}).$$

By dualizing (18) we have $\operatorname{cl}\operatorname{conv} T_\Omega(\bar{z}) = \operatorname{cl}\operatorname{conv} T^{\rm lin}_{P,D}(\bar{z})$ and (33) follows. □

Corollary 1. *Assume that $\bar{z}$ is a sharp minimum for* (1) *and assume that GGCQ is fulfilled, while the mapping $u \rightrightarrows \nabla P(\bar{z})u - T_D(P(\bar{z}))$ is metrically subregular at* $(0, 0)$. *Then there is $w \in T_D(P(\bar{z}))$ and a multiplier $w^* \in \widehat{N}_{T_D(P(\bar{z}))}(w)$ such that $\nabla f(\bar{z}) + \nabla P(\bar{z})^* w^* = 0$.*

Proof. The statement follows immediately from Proposition 1, because by Lemma 7 the second alternative of Proposition 1 is not possible. □

Note that the conclusion of Corollary 1 can also hold in situations when $\bar{z}$ is not a sharp minimum. Besides the cases when there does not exist a direction $\bar{u}$ fulfilling the conditions of the second alternative of Proposition 1, the first alternative of Proposition 1 holds true if there exists some direction $\bar{u}$ satisfying $\langle \nabla f(\bar{z}), \bar{u}\rangle = 0$, $\nabla P(\bar{z})\bar{u} \in T_D(P(\bar{z}))$ such that $\bar{u}$ is an S-stationary solution of (25) because then $0 \in \nabla f(\bar{z}) + \nabla P(\bar{z})^* \widehat{N}_{T_D(P(\bar{z}))}(\nabla P(\bar{z})\bar{u})$ by the definition of S-stationarity. By Theorem 1 we know that the condition

$$\nabla P(\bar{z})\mathbb{R}^d + \mathcal{L}(T_{T_D(P(\bar{z}))})(\nabla P(\bar{z})\bar{u}) = \mathbb{R}^s$$

is sufficient for S-stationarity of $\bar{u}$ and since $\mathcal{L}(T_{T_D(P(\bar{z}))})(\nabla P(\bar{z})\bar{u})$ is always larger than $\mathcal{L}(T_D(P(\bar{z})))$ it is possible that such an S-stationary solution $\bar{u}$ of (25) exists even if $\bar{z}$ is not S-stationary for (1).

We now turn to the proof of Proposition 1. At first we need some prerequisites. As introduced in the recent paper by Benko and Gfrerer [3], consider the program

$$\min_{(u,y)\in\mathbb{R}^d\times\mathbb{R}^s} \langle \nabla f(\bar{z}), u\rangle + \frac{1}{2}\|y\|^2 \quad \text{subject to} \quad \nabla P(\bar{z})u + y \in T_D(P(\bar{z})). \tag{34}$$

Lemma 8. *Assume that the assumptions of Proposition 1 are fulfilled. Then MSCQ holds for the system $\nabla P(\bar{z})u + y \in T_D(P(\bar{z}))$ at every point $(\bar{u}, \bar{y})$ feasible for (34). Further, the program (34) is bounded below and every B-stationary solution $(\bar{u}, \bar{y})$ is also S-stationary, i.e. there is some multiplier $w^* \in \widehat{N}_{T_D(P(\bar{z}))}(\nabla P(\bar{z})\bar{u} + \bar{y})$ such that*

$$\nabla f(\bar{z}) + \nabla P(\bar{z})^* w^* = 0, \quad \bar{y} + w^* = 0. \tag{35}$$

Proof. Consider the set-valued mapping $M(u, y) := \nabla P(\bar{z})u + y - T_D(P(\bar{z}))$. Given any $(u, y) \in \mathbb{R}^d \times \mathbb{R}^s$ we can find $v \in M(u, y)$ such that $\|v\| = \mathrm{d}(0, M(u, y))$ because $M(u, y)$ is closed. Then $0 \in M(u, y - v)$ showing that

$$\mathrm{d}((u, y), M^{-1}(0)) \le \|v\| = \mathrm{d}(0, M(u, y))$$

and MSCQ for the system $\nabla P(\bar{z})u + y \in T_D(P(\bar{z}))$ at every point $(\bar{u}, \bar{y})$ feasible for (34) follows. In order to show the boundedness of the program (34) assume on the contrary that (34) is unbounded below and consider a sequence (u_k, y_k) with $\nabla P(\bar{z})u_k + y_k \in T_D(P(\bar{z}))$ and $\langle \nabla f(\bar{z}), u_k\rangle + \frac{1}{2}\|y_k\|^2 \to -\infty$. Since the mapping $u \rightrightarrows \nabla P(\bar{z})u - T_D(P(\bar{z}))$ is assumed to be metrically subregular and its graph is a closed cone, by Lemma 3 we can find another sequence $\tilde{u}_k$ with $\nabla P(\bar{z})\tilde{u}_k \in T_D(P(\bar{z}))$ and

$$\|\tilde{u}_k - u_k\| \le \kappa \mathrm{d}(\nabla P(\bar{z})u_k, T_D(P(\bar{z}))) \le \kappa\|y_k\|.$$

Because $\bar{z}$ is B-stationary for the program (1) we have $\langle \nabla f(\bar{z}), \tilde{u}_k\rangle \ge 0$, implying

$$\langle \nabla f(\bar{z}), u_k\rangle + \frac{1}{2}\|y_k\|^2 \ge \langle \nabla f(\bar{z}), u_k - \tilde{u}_k\rangle + \frac{1}{2}\|y_k\|^2 \ge -\kappa\|\nabla f(\bar{z})\|\|y_k\| + \frac{1}{2}\|y_k\|^2 \to -\infty,$$

which is obviously not possible. Hence, (34) is bounded below. Finally, the last statement about S-stationarity of B-stationary solutions follows immediately from Theorem 1 applied to (34). □

Lemma 9. *Consider the program*

$$\min_{z\in\mathbb{R}^d} q(z) := \frac{1}{2}z^T B z + b^T z \quad \textit{subject to} \quad Az \in C, \tag{36}$$

where B denotes a positive semidefinite $d\times d$-matrix, $b\in\mathbb{R}^d$, A is an $s\times d$ matrix and $C\subset\mathbb{R}^s$ is a polyhedral set. Then exactly one of the following alternatives can occur:

1. *The program* (36) *is infeasible*
2. *The program* (36) *is unbounded below, i.e. there is a sequence z_k satisfying $Az_k\in C$ and $\lim_{k\to\infty} q(z_k) = -\infty$.*
3. *There exists a global solution $\bar z$.*

Proof. It suffices to show that the program (36) has a global solution if it is feasible and bounded below. Let C be the union of the convex polyhedral sets $C_1,\ldots C_p$ and consider for each i the convex quadratic program

$$\min_z q(z) \quad \text{subject to} \quad Az\in C_i.$$

If this program is feasible, then it must possess a global solution $\bar z_i$, since otherwise by [3, Lemma 4] there would exist a direction w satisfying $Aw\in 0^+C_i$ (the recession cone of C_i), $Bw=0$ and $b^Tw<0$ contradicting the boundedness of (36). Then the one of the $\bar z_i$ who has the samllest objective function value is a global solution of (36). □

Proof of Proposition 1. Assuming that the first condition (21) of Proposition 1 is not fulfilled we will show that the second condition must be fulfilled. If the first condition is not fulfilled, then problem (34) cannot have a global solution, because every global solution $(\bar u,\bar y)$ would be B-stationary and therefore also fulfilling the S-stationary conditions (35) and consequently also the first condition (21) of Proposition 1. On the other hand, the program (34) is bounded below and hence we can find a sequence (u_k,y_k) satisfying $\nabla P(\bar z)u_k + y_k\in T_D(P(\bar z))\ \forall k$ and

$$\lim_{k\to\infty}\langle\nabla f(\bar z),u_k\rangle + \frac{1}{2}\|y_k\|^2 = \gamma := \inf\{\langle\nabla f(\bar z),u\rangle+\frac{1}{2}\|y\|^2 \mid \nabla P(\bar z)u+y\in T_D(P(\bar z))\}. \tag{37}$$

It follows that $\gamma<0$ and without loss of generality we can assume that $\langle\nabla f(\bar z),u_k\rangle<0$ for all k implying $y_k\neq 0$ by B-stationarity of $\bar z$. Next we can assume without loss of generality that u_k is the element u with minimal norm fulfilling $\langle\nabla f(\bar z),u\rangle = \langle\nabla f(\bar z),u_k\rangle, \nabla P(\bar z)u + y_k\in T_D(P(\bar z))$. The sequence u_k must be unbounded because otherwise the sequence y_k must be bounded as well and thus (u_k,y_k) possesses some limit point $(\bar u,\bar y)$ which would be a global solution of (34). Thus by passing to a subsequence we can assume that $\lim_k\|u_k\|=\infty$ and that $u_k/\|u_k\|$ converges to some $\bar u$. From

$$0=\limsup_{k\to\infty}\frac{\gamma}{\|u_k\|^2} = \limsup_{k\to\infty}\Big(\frac{\langle\nabla f(\bar z),u_k\rangle}{\|u_k\|^2}+\frac{\|y_k\|^2}{2\|u_k\|^2}\Big) = \limsup_{k\to\infty}\frac{\|y_k\|^2}{2\|u_k\|^2}$$

we conclude $\|y_k\|/\|u_k\| \to 0$. Hence

$$\langle \nabla f(\bar z), \bar u\rangle = \lim_{k\to\infty} \frac{\langle \nabla f(\bar z), u_k\rangle}{\|u_k\|} \le 0,$$
$$\nabla P(\bar z)\bar u = \lim_{k\to\infty} \frac{1}{\|u_k\|}\big(\nabla P(\bar z)u_k + y_k\big) \in T_D(P(\bar z)),$$

implying $\bar u \in T^{\rm lin}_{P,D}(\bar z)$. Since $\bar z$ is B-stationary for (1) it follows from GGCQ that $\langle \nabla f(\bar z), \bar u\rangle = 0$ and that $\bar u$ is a global solution of the program

$$\min_u \langle \nabla f(\bar z), u\rangle \quad \text{subject to} \quad u \in T^{\rm lin}_{P,D}(\bar z).$$

Hence the B-stationarity condition (24) follows. Next we show (22) by contraposition. Assuming that $\nabla P(\bar z)\bar u \in \mathcal{L}(T_D(P(\bar z)))$, we have $\langle \nabla f(\bar z), u_k - \|u_k\|\bar u\rangle = \langle \nabla f(\bar z), u_k\rangle$ and $\nabla P(\bar z)(u_k - \|u_k\|\bar u) + y_k \in T_D(P(\bar z))$. Since $\|(u_k - \|u_k\|\bar u)\| = \|u_k\|\|\frac{u_k}{\|u_k\|} - \bar u\| < \|u_k\|$ for k sufficiently large, we get a contradiction to our choice of u_k and therefore $\nabla P(\bar z)\bar u \notin \mathcal{L}(T_D(P(\bar z)))$.

There remains to show that $T_D(P(\bar z))$ is not locally polyhedral near $\nabla P(\bar z)\bar u$. Assuming on the contrary that $T_D(P(\bar z))$ is locally polyhedral near $\nabla P(\bar z)\bar u$, we can find a polyhedral set C and a neighborhood W of $\nabla P(\bar z)\bar u$ such that $T_D(P(\bar z)) \cap W = C \cap W$. We can choose the neighborhood W as a convex polyhedral set, e.g. as a sufficiently small ball around $\nabla P(\bar z)\bar u$ with respect to the maximum norm. Hence we can assume that $C\cap W$ is polyhedral and is the union of the convex polyhedral sets $C_1, \ldots, C_q$ having the representations $C_i = \{w \mid \langle a_{ij}, w\rangle \le \alpha_{ij}, j = 1, \ldots, p_i\}$. Consider the set

$$\bigcup_{\beta\ge 1} \beta C_i = \pi(\{(w,\beta) \mid \langle a_{ij}, w\rangle - \beta\alpha_{ij} \le 0, j = 1, \ldots, p_i,\ \beta \ge 1\}),$$

where $\pi(w,\beta) := w$. By [22, Theorem 19.3] this set is a convex polyhedral set, implying that the set

$$\bigcup_{\beta\ge1} \beta(T_D(P(\bar z)) \cap W) = \bigcup_{\beta\ge1} \beta(C \cap W) = \bigcup_{i=1}^{p} \bigcup_{\beta\ge1} \beta C_i$$

is polyhedral. Consider the optimization problem

$$\min_{u,y} \langle \nabla f(\bar z), u\rangle + \frac{1}{2}\|y\|^2 \quad \text{subject to} \quad \nabla P(\bar z)u + y \in \bigcup_{\beta\ge1} \beta(T_D(P(\bar z)) \cap W). \tag{38}$$

Since $\bigcup_{\beta\ge1} \beta(T_D(P(\bar z))\cap W) \subset \bigcup_{\beta\ge1} \beta T_D(P(\bar z))) = T_D(P(\bar z))$, we conclude from Lemma 8 that the problem (38) is bounded below and thus by Lemma 9 it possesses a global solution $(\tilde u, \tilde y)$. By the construction of $\bar u$ we have $(\nabla P(\bar z)u_k + y_k)/\|u_k\| \in C\cap W$ for all k sufficiently large and thus $(\nabla P(\bar z)u_k + y_k) \in \bigcup_{\beta\ge1} \beta(T_D(P(\bar z))\cap W)$. This shows $\langle \nabla f(\bar z), \tilde u\rangle + \frac{1}{2}\|\tilde y\|^2 \le \langle \nabla f(\bar z), u_k\rangle + \frac{1}{2}\|y_k\|^2$ and from (37) we obtain that $(\tilde u, \tilde y)$ is a global solution of (34), a contradiction. Therefore, $T_D(P(\bar z))$ is not locally polyhedral near $\nabla P(\bar z)\bar u$ and this completes the proof. $\square$

For the sake of completeness we state also the following extension of Proposition 1, which exploits some additional features in case of problems of the form (2). Rewriting this problem in the form (1), the set D is the graph of Q and then the tangent cone to D is the graph of another multifunction, the so-called graphical derivative.

Proposition 2. *In addition to the assumptions of Theorem 3 assume that $T_D(P(\bar z))$ is the graph of a set-valued mapping $M = M_c + M_p$, where $M_c, M_p : \mathbb{R}^r \rightrightarrows \mathbb{R}^{s-r}$ are set-valued mappings whose graphs are closed cones, M_p is polyhedral and there is some real C such that*

$$\|t\| \le C\|w\| \ \forall (w,t) \in \operatorname{gph} M_c. \tag{39}$$

Then either there is $w \in T_D(P(\bar z))$ and a multiplier $w^* \in \widehat{N}_{T_D(P(\bar z))}(w)$ fulfilling (21) or there is some $\bar u \in T^{\rm lin}_{P,D}(\bar z)$ fulfilling (22),(23) and (24) such that $T_D(P(\bar z))$ is not locally polyhedral near $\nabla P(\bar z)\bar u$ and there is some $\bar w \neq 0$ with

$$\nabla P(\bar z)\bar u \in \{\bar w\} \times M(\bar w). \tag{40}$$

Proof. We only have to show (40) and we can proceed quite similar as in the proof of Proposition 1. Assuming that we cannot fulfill (21), let (u_k, y_k) denote a sequence satisfying $\nabla P(\bar z)u_k + y_k \in T_D(P(\bar z))$ and (37). Let w_k and $t_k \in M_c(w_k)$ be given by $\nabla P(\bar z)u_k + y_k \in \{w_k\} \times (t_k + M_p(w_k))$ and consider for each k the problem

$$\min \langle \nabla f(\bar z), u\rangle + \frac{1}{2}\|y\|^2 \text{ subject to } \nabla P(\bar z)u + y \in (w_k, t_k + M_p(w_k)). \tag{41}$$

Since $M_p(w_k)$ is a polyhedral set, by Lemma 9 this problem has a global solution and we now claim that there is also a global solution $(\tilde u_k, \tilde y_k)$ fulfilling

$$\|(\tilde u_k, \tilde y_k)\| \le \gamma_1 + \gamma_2(\|t_k\| + \|w_k\|),$$

where γ_1, γ_2 do not depend on k. Indeed, let $\operatorname{gph} M_p$ be the union of the convex polyhedral sets C_i, $i = 1, \ldots, p$ with representation

$$C_i = \{(w, t_p) \,|\, \langle a_{ij}, w\rangle + \langle b_{ij}, t_p\rangle \le \alpha_{ij},\ j = 1, \ldots, q_i\}$$

and consider for each i and each index set $J \subset \{1, \ldots, q_i\}$ the set $S(i, J, w_k, t_c)$ consisting of all $(u, y, t_p, \mu_1, \mu_2, \lambda) \in \mathbb{R}^d \times \mathbb{R}^s \times \mathbb{R}^{s-r} \times \mathbb{R}^r \times \mathbb{R}^{s-r} \times \mathbb{R}^{q_i}$ satisfying the system of linear equalities and linear inequalities

$$\nabla P(\bar z)^* \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = -\nabla f(\bar z),\ y + \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = 0 \tag{42}$$

$$-\mu_2 + \sum_{j\in J} \lambda_i b_{ij} = 0, \lambda_j \ge 0,\ j \in J, \lambda_j = 0,\ j \in \{1, \ldots, q_i\} \setminus J \tag{43}$$

$$\nabla P(\bar z)u + y - (0, t_p) = (w_k, t_c) \tag{44}$$

$$\langle b_{ij}, t_p\rangle \begin{cases} = \alpha_{ij} - \langle a_{ij}, w_k\rangle & \text{if } j \in J, \\ \le \alpha_{ij} - \langle a_{ij}, w_k\rangle & \text{if } j \notin J. \end{cases} \tag{45}$$

By Hoffman's error bound there is some constant $\gamma^{i,J}$ such that

$$\mathrm{d}(0, S(i,J,w_k,t_c)) \le \gamma^{i,J}(\|\nabla f(\bar x)\| + \|w_k\| + \|t_c\| + \sum_{j=1}^{q_i} |\alpha_{ij} - \langle a_{ij}, w_k\rangle|$$

whenever $S(i,J,w_k,t_c) \neq \emptyset$. Note that for every $(u,y,t_p,\mu_1,\mu_2,\lambda) \in S(i,J,w_k,t_c)$ the triple (u,y,t_p) is a global solution of the convex quadratic program

$$\min \langle \nabla f(\bar z), u\rangle + \frac{1}{2}\|y\|^2 \text{ subject to } \nabla P(\bar z)u + y - (0,t_p) = (w_k,t_c), (w_k,t_p) \in C_i \tag{46}$$

because the equations (42)-(45) constitute the Karush-Kuhn-Tucker conditions for this problem. Conversely, for every solution of (u,y,t_p) of this program there must exist multipliers (μ_1,μ_2,λ) such that $(u,y,t_p,\mu_1,\mu_2,\lambda)$ fulfills the Karush-Kuhn-Tucker conditions and thus $(u,y,t_p,\mu_1,\mu_2,\lambda) \in S(i,J,w_k,t_c)$ with $J := \{j \mid \lambda_j > 0\}$.

Now let (u,y) denote a global solution of (41) and let $t_p \in M_p(w_k)$ be given by $\nabla P(\bar z)u + y - (0,t_p) = (w_k,t_c)$. Consider i such that $(w_k,t_p) \in C_i$. Then the triple (u,y,t_p) is a global solution of (46) and we can find some index set J such that $S(i,J,w_k,t_c) \neq \emptyset$. Obviously this set is closed and thus we can find $(\tilde u, \tilde y, \tilde t_p, \tilde\mu_1, \tilde\mu_2) \in S(i,J,w_k,t_c)$ such that $\|(\tilde u, \tilde y, \tilde t_p, \tilde\mu_1, \tilde\mu_2)\| = \mathrm{d}(0, S(i,J,w_k,t_c))$, implying

$$\begin{aligned}\|(\tilde u,\tilde y)\| &\le \|(\tilde u, \tilde y, \tilde t_p, \tilde\mu_1, \tilde\mu_2)\| \le \gamma^{i,J}(\|\nabla f(\bar x)\| + \|w_k\| + \|t_c\| + \sum_{j=1}^{q_i} |\alpha_{ij} - \langle a_{ij}, w_k\rangle|) \\ &\le \gamma^{i,J}(\|\nabla f(\bar x)\| + \sum_{j=1}^{q_i}|\alpha_{ij}|) + \gamma^{i,J}(\|t_c\| + (1 + \sum_{j=1}^{q_i}\|a_{ij}\|)\|w_k\|.\end{aligned}$$

Since both $(\tilde u, \tilde y, \tilde t_p)$ and (u,y,t_p) constitute global solutions of (46) and (u,y) is a global solution of (41), $(\tilde u, \tilde y)$ is a global solution of (41) and our claim follows with $(\tilde u_k, \tilde y_k) = (\tilde u, \tilde y)$ and

$$\gamma_1 = \max_{i,J} \gamma^{i,J}(\|\nabla f(\bar x)\| + \sum_{j=1}^{q_i}|\alpha_{ij}|),\ \gamma_2 = \max_{i,J}\gamma^{i,J}(1 + \sum_{j=1}^{q_i}\|a_{ij}\|).$$

Together with (39) we obtain

$$\|(\tilde u_k, \tilde y_k)\| \le \gamma_1 + \gamma_2(1+C)\|w_k\|. \tag{47}$$

Since (u_k,y_k) is feasible for the problem (41), we have $\langle \nabla f(\bar z), \tilde u_k\rangle + \frac{1}{2}\|\tilde y_k\|^2 \le \langle \nabla f(\bar z), u_k\rangle + \frac{1}{2}\|y_k\|^2$ and thus $(\tilde u_k, \tilde y_k)$ is another sequence fulfilling (37). We can proceed as in the proof of Proposition 1 to show that, after passing to a subsequence, the sequence $\tilde u_k/\|\tilde u_k\|$ converges to some $\bar u \in T^{\mathrm{lin}}_{P,D}(\bar z)$ fulfilling (22),(23) and (24) and $T_D(P(\bar z))$ is not locally polyhedral near $\nabla P(\bar z)\bar u$. Because $\nabla P(\bar z)\bar u = \lim_{k\to\infty}(\nabla P(\bar z)\tilde u_k + \tilde y_k)/\|\tilde u_k\|$ and

$$(\nabla P(\bar z)\tilde u_k + \tilde y_k)/\|\tilde u_k\| \in \frac{1}{\|\tilde u_k\|}\Big(\{w_k\} \times M(w_k)\Big) = \{\frac{w_k}{\|\tilde u_k\|}\} \times M\Big(\frac{w_k}{\|\tilde u_k\|}\Big)$$

we conclude that $\frac{w_k}{\|\tilde u_k\|}$ converges to some $\bar w$ such that $\nabla P(\bar z)\bar u \in \{\bar w\} \times M(\bar w)$. From (47) we obtain $1 \le \gamma_2(1+C)\|\bar w\|$ implying $\|\bar w\| > 0$. This completes the proof. □

5 Application to MPEC

In this section we want to demonstrate that the linearized M-stationarity conditions can be applied to the MPEC (7) when it is impossible to compute the limiting normal cone effectively. Recall that this program is given by

$$
\begin{aligned}
\text{(MPEC')} \qquad & \min_{x,y} F(x,y) \\
& \text{s.t. } \hat{P}(x,y) := \begin{pmatrix} (y, -\phi(x,y)) \\ G(x,y) \end{pmatrix} \in \operatorname{gph} \widehat{N}_\Gamma \times \mathbb{R}^p_- =: \hat{D},
\end{aligned}
$$

where $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $\phi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ and $G : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$ are continuously differentiable and $\Gamma := \{y \,|\, g(y) \le 0\}$ is given by a C^2-mapping $g : \mathbb{R}^m \to \mathbb{R}^q$.

For the rest of the section let $(\bar{x}, \bar{y})$ denote a B-stationary solution for the program (MPEC') such that the following assumption is fulfilled:

Assumption 1. *1. MSCQ holds for the lower level system $g(y) \in \mathbb{R}^q_-$ at $\bar{y}$.*

2. GGCQ holds at $(\bar{x}, \bar{y})$ and the mapping

$$
(u, v) \ \rightrightarrows \ \nabla \hat{P}(\bar{x}, \bar{y})(u, v) - T_{\hat{D}}(\hat{P}(\bar{x}, \bar{y}))
$$

is metrically subregular at $((0,0),0)$.

Note that by Remark 1 the second part of Assumption 1 is fulfilled if MSCQ holds for the system $\hat{P}(x,y) \in \hat{D}$ at $(\bar{x}, \bar{y})$. A point-based sufficient condition for the validity of MSCQ for this system is given by [14, Theorem 5].

We need some more notation. We set $\bar{y}^* := -\phi(\bar{x}, \bar{y})$ and denote by

$$
\bar{K}_\Gamma := \mathcal{K}_\Gamma(\bar{y}, \bar{y}^*)
$$

the critical cone for Γ at $(\bar{y}, \bar{y}^*)$. Further we define the *multiplier set*

$$
\bar{\Lambda} := \{\lambda \in N_{\mathbb{R}^q_-}(g(\bar{y})) \,|\, \nabla g(\bar{y})^* \lambda = \bar{y}^*\}
$$

and for every $v \in \bar{K}_\Gamma$ the *directional multiplier set*

$$
\bar{\Lambda}(v) := \arg\max\{v^T \nabla^2 (\lambda^T g)(\bar{y}) v \,|\, \lambda \in \bar{\Lambda}\}.
$$

By [11, Proposition 4.3(iii)] we have $\bar{\Lambda}(v) \neq \emptyset \ \forall v \in \bar{K}_\Gamma$ thanks to Assumption 1(1).

By [14, Proposition 1] we have

$$
T_{\hat{D}}(\hat{P}(\bar{x}, \bar{y})) = T_{\operatorname{gph} \widehat{N}_\Gamma}(\bar{y}, \bar{y}^*) \times T_{\mathbb{R}^p_-}(G(\bar{x}, \bar{y})).
$$

In order to compute the tangent cone $T_{\operatorname{gph} \widehat{N}_\Gamma}(\bar{y}, \bar{y}^*)$ we use the following theorem:

Theorem 4 (cf. [14, Theorem 4]). *Assume that MSCQ holds at $\bar{y}$ for the system $g(y) \in \mathbb{R}^q_-$. Then there is a real $\kappa > 0$ such that the tangent cone to the graph of $\widehat{N}_\Gamma$ at $(\bar{y}, \bar{y}^*)$ can be calculated by*

$$\begin{aligned} & T_{\operatorname{gph} \widehat{N}_\Gamma}(\bar{y}, \bar{y}^*) \\ &= \big\{(v, v^*) \in \mathbb{R}^{2m} \mid \exists\, \lambda \in \bar{\Lambda}(v) \ \ \textit{with} \ \ v^* \in \nabla^2(\lambda^T g)(\bar{y})v + N_{\bar{K}_\Gamma}(v)\big\} \\ &= \big\{(v, v^*) \in \mathbb{R}^{2m} \mid \exists\, \lambda \in \bar{\Lambda}(v) \cap \kappa\|\bar{y}^*\|\mathcal{B}_{\mathbb{R}^q} \ \ \textit{with} \ \ v^* \in \nabla^2(\lambda^T g)(\bar{y})v + N_{\bar{K}_\Gamma}(v)\big\}. \end{aligned} \tag{48}$$

We see that the tangent cone $T_{\hat{D}}(\hat{P}(\bar{x}, \bar{y}))$ is the graph of the multifunction $M(v) = M_c(v) + M_p(v)$, where

$$M_p(v) := N_{\bar{K}_\Gamma}(v) \times T_{\mathbb{R}^p_-}(G(\bar{x}, \bar{y}))$$

is a polyhedral multifunction and

$$M_c(v) := \{\nabla^2(\lambda^T g)(\bar{y})v \mid \lambda \in \bar{\Lambda}(v) \cap \kappa\|\bar{y}^*\|\mathcal{B}_{\mathbb{R}^q}\} \times \{0\}.$$

fulfills (39).

Proposition 3. *Let a critical direction $\bar{v} \in \bar{K}_\Gamma$ be given. If there is an open neighborhood V of $\bar{v}$ and a set $\tilde{\Lambda} \subset \bar{\Lambda}$ such that*

$$\bar{\Lambda}(v) = \tilde{\Lambda} \ \forall v \in (\bar{K}_\Gamma \setminus \{\bar{v}\}) \cap V \tag{49}$$

then

$$T_{\operatorname{gph} \widehat{N}_\Gamma}(\bar{y}, \bar{y}^*) \cap (V \times \mathbb{R}^m) = \{\big(v, \nabla^2(\tilde{\lambda}^T g)(\bar{y})v + z^*\big) \mid z^* \in N_{\bar{K}_\Gamma}(v)\} \cap (V \times \mathbb{R}^m), \tag{50}$$

where $\tilde{\lambda} \in \tilde{\Lambda}$ is an arbitrarily fixed multiplier. In particular, $T_{\operatorname{gph} \widehat{N}_\Gamma}(\bar{y}, \bar{y}^*)$ is locally polyhedral near $(\bar{v}, \bar{v}^*)$ for every $\bar{v}^*$ satisfying $(\bar{v}, \bar{v}^*) \in T_{\operatorname{gph} \widehat{N}_\Gamma}(\bar{y}, \bar{y}^*)$ and

$$\widehat{N}_{T_{\operatorname{gph} \widehat{N}_\Gamma}(\bar{y}, \bar{y}^*)}(\bar{v}, \bar{v}^*) = \big\{(w^*, w) \,|\, (w^* + \nabla^2(\tilde{\lambda}^T g)(\bar{y})w, w) \in \big(\mathcal{K}_{\bar{K}_\Gamma}(\bar{v}, \bar{z}^*)\big)^* \times \mathcal{K}_{\bar{K}_\Gamma}(\bar{v}, \bar{z}^*)\big\}, \tag{51}$$

where $\bar{z}^* := \bar{v}^* - \nabla^2(\tilde{\lambda}^T g)(\bar{y})\bar{v}$.

Proof. Let $\tilde{\lambda} \in \tilde{\Lambda}$ be arbitrarily fixed. We claim that for every $v \in (\bar{K}_\Gamma \setminus \{\bar{v}\}) \cap V$ we have

$$\big\{\nabla^2(\lambda^T g)(y)v \mid \lambda \in \bar{\Lambda}(v)\big\} + N_{\bar{K}_\Gamma}(v) = \nabla^2(\tilde{\lambda}^T g)(y)v + N_{\bar{K}_\Gamma}(v) \tag{52}$$

Indeed, consider $v^* = \nabla^2(\lambda^T g)(y)v + z^*$ with $\lambda \in \bar{\Lambda}(v)$ and $z^* \in N_{\bar{K}_\Gamma}(v)$. Since $\bar{K}_\Gamma$ is a convex polyhedral set, for every $w \in T_{\bar{K}_\Gamma}(v)$ we have $v + \alpha w \in (\bar{K}_\Gamma \setminus \{\bar{v}\}) \cap V$ for all $\alpha \geq 0$ sufficiently small and therefore $(v+\alpha w)^T \nabla^2(\lambda^T g)(\bar{y})(v+\alpha w) = (v+\alpha w)^T \nabla^2(\tilde{\lambda}^T g)(\bar{y})(v+\alpha w)$. Because we also have $v^T \nabla^2(\lambda^T g)(\bar{y})v = v^T \nabla^2(\tilde{\lambda}^T g)(\bar{y})v$ we conclude $v^T \nabla^2\big((\lambda - \tilde{\lambda})^T g\big)(\bar{y})w = 0 \ \forall w \in T_{\bar{K}_\Gamma}(v)$ and consequently $\nabla^2\big((\lambda - \tilde{\lambda})^T g\big)(\bar{y})v \in \mathcal{L}(N_{\bar{K}_\Gamma}(v))$. Thus

$$\begin{aligned} v^* &= \nabla^2(\tilde{\lambda}^T g)(y)v + \nabla^2\big((\lambda - \tilde{\lambda})^T g\big)(\bar{y})v + z^* \\ &\in \nabla^2(\tilde{\lambda}^T g)(y)v + \mathcal{L}(N_{\bar{K}_\Gamma}(v)) + N_{\bar{K}_\Gamma}(v) = \nabla^2(\tilde{\lambda}^T g)(y)v + N_{\bar{K}_\Gamma}(v) \end{aligned}$$

and

$$\left\{\nabla^2(\lambda^T g)(y)v \,|\, \lambda \in \bar\Lambda(v)\right\} + N_{\bar K_\Gamma}(v) \subset \nabla^2(\tilde\lambda^T g)(y)v + N_{\bar K_\Gamma}(v)$$

follows. Since the reverse inclusion obviously holds, our claim (52) is verified. We next show that (52) holds for $v = \bar v$ as well. If $\bar v = 0$ then (52) obviously holds for $v = \bar v$. On the other hand, if $\bar v \neq 0$, we can find some $\alpha \neq 1$ sufficiently close to 1 such that $\alpha\bar v \in (\bar K_\Gamma \setminus \{\bar v\}) \cap V$, implying

$$\begin{aligned}
&\alpha\Big(\left\{\nabla^2(\lambda^T g)(y)\bar v \,|\, \lambda \in \bar\Lambda(\bar v)\right\} + N_{\bar K_\Gamma}(\bar v)\Big) = \left\{\nabla^2(\lambda^T g)(y)\alpha\bar v \,|\, \lambda \in \bar\Lambda(\alpha\bar v)\right\} + N_{\bar K_\Gamma}(\alpha\bar v)\\
&= \nabla^2(\tilde\lambda^T g)(y)\alpha\bar v + N_{\bar K_\Gamma}(\alpha\bar v) = \alpha\Big(\nabla^2(\tilde\lambda^T g)(y)\bar v + N_{\bar K_\Gamma}(\bar v)\Big),
\end{aligned}$$

where we have used the relations $\bar\Lambda(\alpha\bar v) = \bar\Lambda(\bar v)$ and $N_{\bar K_\Gamma}(\bar v) = N_{\bar K_\Gamma}(\alpha\bar v) = \alpha N_{\bar K_\Gamma}(\bar v)$. Thus (52) holds for all $v \in \bar K_\Gamma \cap V$ and the representation (50) follows from (48). Since the graph of the normal cone mapping to a convex polyhedral set is a polyhedral set [23], $\operatorname{gph} N_{\bar K_\Gamma}$ is the union of polyhedral convex sets $C_1, \ldots, C_l \subset \mathbb{R}^m \times \mathbb{R}^m$. By taking into account [22, Theorem 19.3] we obtain that $\{(v, \nabla^2(\tilde\lambda^T g)(\bar y)v + z^*) \,|\, z^* \in N_{\bar K_\Gamma}(v)\}$ is the union of the polyhedral convex sets $\{(v, \nabla^2(\tilde\lambda^T g)(\bar y)v + z^*) \,|\, (v, z^*) \in C_i\}$, $i = 1, \ldots, l$. Now it follows from (50) that $T_{\operatorname{gph} \widehat N_\Gamma}(\bar y, \bar y^*)$ is locally polyhedral near $(\bar v, \bar v^*)$ for every $\bar v^*$ satisfying $(\bar v, \bar v^*) \in T_{\operatorname{gph} \widehat N_\Gamma}(\bar y, \bar y^*)$.

By virtue of (50), for every pair $(v, v^*) \in T_{\operatorname{gph} \widehat N_\Gamma}(\bar y, \bar y^*)$ close to $(\bar v, \bar v^*)$ there is a unique element $z^* \in N_{\bar K_\Gamma}(v)$ with $v^* = \nabla^2(\tilde\lambda^T g)(\bar y)v + z^*$. Thus

$$\begin{aligned}
&(w^*, w) \in \widehat N_{T_{\operatorname{gph} \widehat N_\Gamma}(\bar y, \bar y^*)}(\bar v, v^*) \iff \limsup_{(v,v^*) \overset{T_{\operatorname{gph} \widehat N_\Gamma}(\bar y,\bar y^*)}{\longrightarrow} (\bar v, \bar v^*)} \frac{\langle w^*, v - \bar v\rangle + \langle w, v^* - \bar v^*\rangle}{\|(v, v^*) - (\bar v, \bar v^*)\|} \le 0\\
&\iff \limsup_{(v,z^*) \overset{\operatorname{gph} N_{\bar K_\Gamma}}{\longrightarrow} (\bar v, \bar z^*)} \frac{\langle w^*, v - \bar v\rangle + \langle w, \nabla^2(\tilde\lambda^T g)(\bar y)v + z^* - \nabla^2(\tilde\lambda^T g)(\bar y)\bar v - \bar z^*\rangle}{\|(v, \nabla^2(\tilde\lambda^T g)(\bar y)v + z^*) - (\bar v, \nabla^2(\tilde\lambda^T g)(\bar y)\bar v + \bar z^*)\|} \le 0\\
&\iff \limsup_{(v,z^*) \overset{\operatorname{gph} N_{\bar K_\Gamma}}{\longrightarrow} (\bar v, \bar z^*)} \frac{\langle w^* + \nabla^2(\tilde\lambda^T g)(\bar y)w, v - \bar v\rangle + \langle w, z^* - \bar z^*\rangle}{\|(v, z^*) - (\bar v, \bar z^*)\|} \le 0\\
&\iff (w^* + \nabla^2(\tilde\lambda^T g)(\bar y)w, w) \in \widehat N_{\operatorname{gph} N_{\bar K_\Gamma}}(\bar v, \bar z^*)
\end{aligned}$$

and (51) follows from the identity $\widehat N_{\operatorname{gph} N_{\bar K_\Gamma}}(\bar v, \bar z^*) = \big(\mathcal{K}_{\bar K_\Gamma}(\bar v, \bar z^*)\big)^* \times \mathcal{K}_{\bar K_\Gamma}(\bar v, \bar z^*)$, cf. [6, Equation (13)]. □

We are now in the position to state the main result of this section.

Theorem 5. *Assume that $(\bar x, \bar y)$ is B-stationary for the program (MPEC'), assume that Assumption 1 is fulfilled and that there is a set $\tilde\Lambda \subset \bar\Lambda$ such that*

$$\bar\Lambda(v) = \tilde\Lambda \;\forall v \in \bar K_\Gamma \setminus \{0\}. \tag{53}$$

Then for every $\tilde{\lambda} \in \tilde{\Lambda}$ there are $v \in \bar{K}_\Gamma$, $z^* \in N_{\bar{K}_\Gamma}(v)$ and multipliers $w \in \mathcal{K}_{\bar{K}_\Gamma}(v, \bar{z}^*)$, $\mu \in N_{\mathbb{R}^p_-}(G(\bar{x}, \bar{y}))$ such that

$$
\begin{aligned}
&0 = \nabla_x F(\bar{x}, \bar{y}) - \nabla_x \phi(\bar{x}, \bar{y})^* w + \nabla_x G(\bar{x}, \bar{y})^* \mu \\
&0 \in \nabla_y F(\bar{x}, \bar{y}) - \nabla^2 (\tilde{\lambda}^T g)(\bar{y}) w - \nabla_y \phi(\bar{x}, \bar{y})^* w + \nabla_y G(\bar{x}, \bar{y})^* \mu + \big(\mathcal{K}_{\bar{K}_\Gamma}(v, \bar{z}^*)\big)^*.
\end{aligned}
$$

Proof. By (53) and Proposition 3 we obtain that $T_{\mathrm{gph}\, \widehat{N}_\Gamma}(\bar{y}, \bar{y}^*)$ is locally polyhedral near every $(v, v^*) \in T_{\mathrm{gph}\, \widehat{N}_\Gamma}(\bar{y}, \bar{y}^*)$. Since $T_{\mathbb{R}^p_-}(G(\bar{x}, \bar{y}))$ is a convex polyhedral set, $T_{\hat{D}}(\bar{y}, \bar{y}^*, G(\bar{x}, \bar{y}))$ is polyhedral near every direction $(v, v^*, t) \in T_{\hat{D}}(\bar{y}, \bar{y}^*, G(\bar{x}, \bar{y}))$. Hence, by Proposition 1 there exists a direction $(v, v^*, t) \in T_{\hat{D}}(\bar{y}, \bar{y}^*, G(\bar{x}, \bar{y}))$ and a regular normal $(w^*, w, \mu) \in \widehat{N}_{T_{\hat{D}}(\bar{y}, \bar{y}^*, G(\bar{x}, \bar{y}))}(v, v^*, t) = \widehat{N}_{T_{\mathrm{gph}\, \widehat{N}_\Gamma}(\bar{y}, \bar{y}^*)}(v, v^*) \times N_{T_{\mathbb{R}^p_-}(G(\bar{x}, \bar{y}))}(t)$ such that

$$
0 = \nabla F(\bar{x}, \bar{y}) + \nabla \hat{P}(\bar{x}, \bar{y})^* \begin{pmatrix} w^* \\ w \\ \mu \end{pmatrix} = \begin{pmatrix} \nabla_x F(\bar{x}, \bar{y}) - \nabla_x \phi(\bar{x}, \bar{y})^* w + \nabla_x G(\bar{x}, \bar{y})^* \mu \\ \nabla_y F(\bar{x}, \bar{y}) + w^* - \nabla_y \phi(\bar{x}, \bar{y})^* w + \nabla_y G(\bar{x}, \bar{y})^* \mu \end{pmatrix}
$$

By utilizing (51) and the well-known identity $N_{\mathbb{R}^p_-}(G(\bar{x}, \bar{y})) = \bigcup_{t \in T_{\mathbb{R}^p_-}(G(\bar{x}, \bar{y}))} N_{T_{\mathbb{R}^p_-}(G(\bar{x}, \bar{y}))}(t)$ the assertion follows. $\square$

Recall that the inequalities $g(y) \leq 0$ satisfy the *constant rank constraint qualification* (CRCQ) at a feasible point $\bar{y}$ if for each subset $I \subseteq \{i \in \{1, \ldots, q\} \,|\, g_i(\bar{y}) = 0\}$ there is a neighborhood V of $\bar{y}$ such that the rank of $\{\nabla g_i(y) \,|\, i \in I\}$ is a constant value on V. It was shown in [11, Proposition 5.3] that CRCQ at $\bar{y}$ is a sufficient condition for (53) to hold. By applying [11, Proposition 5.3] to the system $\tilde{g}(y) \leq 0$, where

$$
\tilde{g}_i(y) = g_i(\bar{y}) + \langle \nabla g_i(\bar{y}), y - \bar{y} \rangle + \frac{1}{2}(y - \bar{y})^* \nabla^2 g_i(\bar{y})(y - \bar{y}),\ i = 1, \ldots, q
$$

it follows that it is sufficient to require CRCQ for the system $\tilde{g}(y) \leq 0$ in order to guarantee (53). However, it is easy to find examples where the condition (53) is fulfilled but CRCQ neither for the system $g(y) \leq 0$ nor the system $\tilde{g}(y) \leq 0$ holds.

The following example demonstrates the benefit of the necessary optimality conditions of Theorem 5

Example 1. *Consider the problem*

$$
\min_{x \in \mathbb{R}, y \in \mathbb{R}^3} x - 2y_3 \quad \text{subject to} \quad 0 \in (y_1, y_2, -x + y_3) + \widehat{N}_\Gamma(y)
$$

with

$$
\Gamma := \left\{ y \in \mathbb{R}^3 \,|\, \begin{array}{l} g_1(y) := y_3 - y_1^3 \leq 0 \\ g_2(y) := y_3 - a^3 y_2^3 \leq 0 \end{array} \right\},
$$

where $a > 0$ denotes a fixed parameter. Then $\bar{x} = 0$, $\bar{y} = (0, 0, 0)$ is a local solution. Obviously MFCQ is fulfilled at $\bar{y}$ and straightforward calculations yield $\bar{y}^* = (0, 0, 0)$, $\bar{K}_\Gamma = \mathbb{R} \times \mathbb{R} \times \mathbb{R}_-$ and

$$
\bar{\Lambda} = \bar{\Lambda}(v) = \{(0, 0)\}\ \forall v \in \bar{K}_\Gamma.
$$

Thus condition (53) *is fulfilled and the first-order optimality condition of Theorem 5 must hold. Indeed, taking* $\tilde\lambda = (0,0)$, $v = z^* = (0,0,0)$ *we have* $\mathcal{K}_{\bar K_\Gamma}(v,z^*) = \bar K_\Gamma$, $\big(\mathcal{K}_{\bar K_\Gamma}(v,z^*)\big)^* = \{0\}\times\{0\}\times\mathbb{R}_+$ *and with* $w = (0,0,-1)$ *we obtain*

$$\begin{aligned}
&\nabla_x F(\bar x,\bar y) - \nabla_x\phi(\bar x,\bar y)^* w = 1 - (0\ 0\ -1)\begin{pmatrix}0\\0\\-1\end{pmatrix} = 0,\\
&-\big(\nabla_y F(\bar x,\bar y) - \nabla^2(\tilde\lambda^T g)(\bar y)w - \nabla_y\phi(\bar x,\bar y)^* w\big)\\
&= -\left(\begin{pmatrix}0\\0\\-2\end{pmatrix} - \begin{pmatrix}0\\0\\0\end{pmatrix} - \begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}\begin{pmatrix}0\\0\\-1\end{pmatrix}\right) = \begin{pmatrix}0\\0\\1\end{pmatrix} \in \big(\mathcal{K}_{\bar K_\Gamma}(v,\bar z^*)\big)^*
\end{aligned}$$

verifying the first-order optimality conditions of Theorem 5.

In [12, Example 1] the limiting normal cone $N_{\mathrm{gph}\,\Gamma}(\bar y,\bar y^*)$ *was computed explicitly. It appears that it depends on the parameter* a *and thus a point-based representation of the limiting normal cone in terms of first-order and second-order derivatives of* g *is not possible. This shows the difficulty of verifying the M-stationarity conditions at the solution.*

So far we have only considered linearized M-stationarity conditions for the MPEC (7) under the assumption (53), which allows the application of Theorem 3 with $k = 0$. In a forthcoming paper we will formulate the linearized M-stationarity conditions for this problem for the general case. Anticipating the main result of that paper we will show with the help of Proposition 2 that Theorem 3 holds with $k = 1$.

6 Concluding remarks and future research

In this paper we considered new first-order optimality conditions for general optimization problems which are stronger than the commonly used M-stationarity conditions. The key idea is to apply the M-stationarity conditions not to the original problem but to the linearized problem and to repeat this procedure. As a final result we obtain that the multiplier is not only a limiting normal but also a regular normal to tangent cone to a series of tangent cones. Because the optimality conditions are based on a repeated linearization process we use the term *linearized M-stationarity conditions.*

The applicability of the new optimality conditions are demonstrated at the basis of a special MPEC, where the equilibrium is modeled via a general equation involving the normal cone to a set given by C^2-inequalities. Under a certain additional condition we explicitly stated the optimality conditions in terms of the problem data at the reference point. This additional assumption ensures that the linearization process has not to be repeated. We presented an example where the M-stationarity conditions cannot be stated effectively by the difficulty of computing the limiting normal cone, whereas our results fully apply. We plan to drop this additional assumption in a forthcoming paper to obtain the linearized M-stationarity condition for this MPEC in the general case.

A further goal is the application of the developed theory to other problem classes, e.g. to MPECs involving the normal cone to sets appearing in second-order cone programming and semidefinite programming. In particular in the latter case we expect that the linearization process has to be eventually repeated more than once.

Another direction of future research could be the investigation of the sufficiency of the linearized M-stationarity conditions for B-stationarity. Similar as in [26] one could look for properties of the problem functions which ensure that the reference point is a globally or locally optimal solution. Another approach could be the fulfilment of some linearized M-stationary condition in every nonzero critical direction similar to the concept of *extended M-stationarity* used in [10].

Acknowledgements. The research was partially supported by the Austrian Science Fund (FWF) under grant P29190-N32.

References

[1] L. Adam, R. Henrion, J. Outrata, *On M-stationarity conditions in MPECs and the associated qualification conditions*, Math. Program. Series B, 168 (2018), pp. 229–259.

[2] M. Benko, H. Gfrerer, *On estimating the regular normal cone to constraint systems and stationary conditions*, Optimization, 66 (2017), pp. 61–92.

[3] M. Benko, H. Gfrerer, *New verifiable stationarity concepts for a class of mathematical programs with disjunctive constraints*, Optimization, 67 (2018), pp. 1–23.

[4] F. H. Clarke, *Necessary conditions for Nonsmooth Problems in Optimal Control and the Calculus of Variations*, Ph.D. dissertation, University of Washington, Seattle, 1973.

[5] N.H. Chieu, L.V. Hien, *Computation of graphical derivative for a class of normal cone mappings under a very weak condition*, SIAM J. Optim., 27 (2017), pp. 190-204.

[6] A. L. Dontchev, R. T. Rockafellar, *Characterizations of strong regularity for variational inequalities over polyhedral convex sets*, SIAM J. Optim., 6 (1996), pp. 1087–1105.

[7] A.L. Dontchev, R.T. Rockafellar, *Regularity and conditioning of solution mappings in variational analysis*, Set-Valued Var. Anal., 12 (2004), pp. 79-109.

[8] M. L. Flegel, C. Kanzow, J. V. Outrata, *Optimality conditions for disjunctive programs with application to mathematical programs with equilibrium constraints*, Set-Valued Anal., 15 (2007), pp. 139–162.

[9] H. Gfrerer, *First order and second order characterizations of metric subregularity and calmness of constraint set mappings*, SIAM J. Optim., 21 (2011), pp. 1439–1474.

[10] H. Gfrerer, *Optimality conditions for disjunctive programs based on generalized differentiation with application to mathematical programs with equilibrium constraints*, SIAM J. Optim., 24 (2014), pp. 898–931.

[11] H. Gfrerer, B.S. Mordukhovich, *Complete characterizations of tilt stability in nonlinear programming under weakest qualification conditions*, SIAM J. Optim., 25 (2015), pp. 2081–2119.

[12] H. Gfrerer and J. V. Outrata, *On computation of limiting coderivatives of the normal-cone mapping to inequality systems and their applications*, Optimization, 65 (2016), pp. 671–700.

[13] H. Gfrerer, J. V. Outrata, *On computation of generalized derivatives of the normal-cone mapping and their applications*, Math. Oper. Res. 41 (2016), pp. 1535–1556.

[14] H. Gfrerer, J. J. Ye, *New constraint qualifications for mathematical programs with equilibrium constraints via variational analysis*, SIAM J. Optim., 27 (2017), pp. 842–865.

[15] R. Henrion, J. V. Outrata, *Calmness of constraint systems with applications*, Math. Program., Ser. B, 104 (2005), pp. 437–464.

[16] R. Henrion, J. V. Outrata, T. Surowiec, *On the coderivative of normal cone mapping to inequality systems*, Nonlinear Anal., 71 (2009), pp. 1213–1226.

[17] Z.-Q. Luo, J.-S. Pang, D. Ralph, *Mathematical Programs with Equilibrium Constraints*, Cambridge University Press, Cambridge, UK, 1996.

[18] B. S. Mordukhovich, J. V. Outrata, *On second-order subdifferentials and their applications*, SIAM J. Optim., 12 (2001), 139–169.

[19] J. V. Outrata, *Optimality conditions for a class of mathematical programs with equilibrium constraints*, Math. Oper. Res., 24 (1999), pp. 627–644.

[20] S. Scholtes, *Convergence properties of a regularization scheme for mathematical programs with complementarity constraints*, SIAM J. Optim., 11 (2001), pp. 918–936.

[21] H. Scheel, S. Scholtes, *Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity*, Math. Oper. Res., 25 (2000), pp. 1–22.

[22] R. T. Rockafellar, *Convex analysis*, Princeton, New Jersey, 1970.

[23] S. M. Robinson, *Generalized equations and their solutions, part I: basic theory*, in Point-to-set Maps and Mathematical Programming, P. Huard, ed., Mathematical Programming Study, 10, North Holland, Amsterdam, 1979, pp. 128–141.

[24] R.T. Rockafellar, R. J-B. Wets, *Variational analysis*, Springer, Berlin, 1998.

[25] J. J. Ye, *Optimality conditions for optimization problems with complementarity constraints*, SIAM J. Optim., 9 (1999), pp. 374–387.

[26] J. J. Ye, *Necessary and sufficient optimality conditions for mathematical programs with equilibrium constraints*, J. Math. Anal. Appl. 307 (2005), pp. 350–369.