A new problem qualification based on approximate KKT conditions for Lipschitzian optimization with application to bilevel programming

Isabella Käming[a], Andreas Fischer[a] and Alain B. Zemkoho[b]

[a] Institute of Numerical Mathematics, Technische Universität Dresden, 01062 Dresden, Germany
[b] School of Mathematics, University of Southampton, Southampton SO17 1BJ, UK

ABSTRACT
When dealing with general Lipschitzian optimization problems, there are many problem classes where even weak constraint qualifications fail at local minimizers. In contrast to a constraint qualification, a problem qualification does not only rely on the constraints but also on the objective function to guarantee that a local minimizer is a Karush-Kuhn-Tucker (KKT) point. For example, calmness in the sense of Clarke is a problem qualification. In this article, we introduce the Subset Mangasarian-Fromovitz Condition (subMFC). This new problem qualification is derived by means of a nonsmooth version of the approximate KKT conditions, which hold at every local minimizer without further assumptions. A comparison with existing constraint and problem qualifications reveals that subMFC is strictly weaker than quasinormality and can hold even if the local error bound condition, the cone-continuity property, the Guignard constraint qualification and calmness are violated. Furthermore, we emphasize the power of the new problem qualification within the context of bilevel optimization. More precisely, under mild assumptions on the problem data, we suggest a version of subMFC that is tailored to the lower-level value function reformulation. It turns out that this new condition can be satisfied even if the widely used partial calmness condition does not hold.

1. Introduction

We consider mathematical optimization problems of the form

$$\min_x f(x) \quad \text{s.t.} \quad x \in D := \{x \in \mathbb{R}^n \mid g(x) \leq 0\} \tag{P}$$

with locally Lipschitz continuous functions $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^q$. For simplicity, we only regard problems with inequality constraints in the course of this paper. Later on, we will see by examples that equality constraints can be handled as well by splitting them into two inequality constraints. It is known that a local minimizer of problem (P) satisfies the KKT conditions if some suitable constraint qualification (CQ) is fulfilled. Unfortunately, there are situations where even weak CQs fail. In such cases we are faced with the problem that

---

Contact: Isabella Käming. Email: isabella.kaeming@tu-dresden.de

it is hard to analyze whether we can expect the KKT conditions to hold at local minimizers of the problem. For example, this difficulty generically occurs in bilevel optimization [47] or for mathematical programs with complementarity constraints (MPCCs) [5, 37].

Therefore, instead of directly considering the KKT conditions, we are interested in optimality conditions that do not require the fulfillment of a CQ. For the smooth counterpart of problem (P), an established concept is the usage of sequential optimality conditions, as they are valid independently of CQs, see [2, 5, 8, 11]. Moreover, sequential optimality conditions imply the KKT conditions under specially tailored conditions which are weak enough such that they can actually be satisfied [4, 6]. Thus, the procedure of combining sequential optimality conditions with a weak condition is also of interest to obtain optimality conditions for instances of the aforementioned problem classes, and has been carried out in [5, 37] for MPCCs with continuously differentiable problem data. Another advantage of sequential optimality conditions is that they can be used to justify termination criteria for numerical algorithms, as done for augmented Lagrangian methods [11], SQP methods [36], inexact restoration methods [12, 17] and others, like [2, 8, 28].

In this paper, as sequential optimality conditions, we consider the well-known approximate KKT (AKKT) conditions and transfer the definition used in [36] to our nonsmooth setting. The main focus is then to introduce and analyze a new condition which guarantees that a point fulfilling the AKKT conditions is already a KKT point. One could therefore be tempted to assume that, in terms of [11], this new condition belongs to the class of *strict* CQs. Strict CQs are all CQs which ensure that an AKKT point is also a KKT point. For example, the class of strict CQs contains CQs like the linear independence CQ (LICQ), Mangasarian-Fromovitz CQ (MFCQ), constant rank CQ (CRCQ), constant positive linear dependence condition (CPLD) and the cone-continuity property (CCP) [6]. The latter is the weakest strict CQ [6, 11] and also called asymptotic regularity in a nonsmooth context [31]. In contrast to this, the new condition we are interested in does *not* belong to the class of strict CQs. In fact, it directly depends on the specific sequence fulfilling the AKKT conditions for the given point, which is a significant difference to strict CQs that are stated completely independent of AKKT conditions. Due to the latter circumstance, the new condition is *not even* a CQ, as the AKKT conditions and therefore the condition itself involve the objective function. Hence, we refer to such conditions as *problem qualifications*. Similarly to CQs, problem qualifications also guarantee that the KKT conditions hold at a local minimizer. A well-known problem qualification is calmness in the sense of Clarke [13, Definition 6.4.1].

It turns out that the proposed novel coupling of the AKKT conditions with a directly corresponding condition provides a very powerful problem qualification for (P), which we call *Subset Mangasarian-Fromovitz Condition (subMFC)*. Indeed, exploiting the information obtained by the specific fulfilled AKKT conditions enables subMFC to still maintain the most important property of strict CQs, which is serving as a condition to ensure that an AKKT point is already a KKT point. At the same time, it can be shown that subMFC can hold even if CCP (the weakest strict CQ) fails. Besides this advantage, subMFC has the following additional benefits:

- It is easy to understand, since it can be interpreted as a weakened version of the MFCQ.
- A comparison with existing CQs for (P) reveals that subMFC is a quite weak condition that can be satisfied by a wide range of optimization problems. In particular, subMFC is strictly implied by quasinormality. Additionally, examples show that subMFC is independent of the local error bound condition and the Guignard CQ, and that subMFC can hold even if the problem qualification calmness in the sense of Clarke is violated.

- It can be handled easier than CCP, which is hard to verify [6] since it depends on *all* subgradients that occur in the neighborhood of the considered point. With subMFC, generally significantly fewer subgradients need to be considered.
- SubMFC can be used to treat all problem classes covered by (P), which includes MPCCs [27], mathematical programs with vanishing constraints [1] or mathematical programs with switching constraints [30].

As for the last advantage, in this paper, we specifically demonstrate the power of our approach by applying it to bilevel optimization problems. For this purpose, let us consider the standard optimistic bilevel optimization problem

$$\min_{x,y} F(x,y) \quad \text{s.t.} \quad G(x,y) \le 0, \quad y \in S(x) \tag{BP}$$

with the lower-level solution set

$$S(x) \coloneqq \operatorname*{argmin}_y \{\hat{f}(x,y) \mid \hat{g}(x,y) \le 0\},$$

where the functions $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $G : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$, $\hat{f} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and $\hat{g} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^r$ are assumed to be locally Lipschitz continuous. A widely used single-level reformulation of (BP) is based on the lower-level value function $\varphi : \mathbb{R}^n \to \mathbb{R}$, which is defined by

$$\varphi(x) \coloneqq \min_y \{\hat{f}(x,y) \mid \hat{g}(x,y) \le 0\}.$$

To ensure that φ is well-defined, we assume throughout that the lower-level solution set $S(x)$ is not empty for each $x \in \mathbb{R}^n$. The lower-level value function reformulation is then given as

$$\min_{x,y} F(x,y) \quad \text{s.t.} \quad G(x,y) \le 0, \quad \hat{g}(x,y) \le 0, \quad \hat{f}(x,y) - \varphi(x) \le 0. \tag{LLVFR}$$

By definition of φ, this problem is equivalent to (BP) concerning global and local solutions. If φ is locally Lipschitz continuous, it is clear that (LLVFR) is a special case of (P). Thus, we will apply our results related to subMFC to derive the new problem qualification *LLVFR-subMFC* that is tailored to (LLVFR). The major advantages of LLVFR-subMFC are:

- LLVFR-subMFC can hold even if partial calmness is violated. Partial calmness is a condition which has emerged as a standard condition for deriving KKT-type optimality conditions for bilevel optimization problems treated by the lower-level value function reformulation [14–16, 19, 47].
- In contrast to most of the existing literature, we do not require any additional assumptions on the bilevel problem besides the local Lipschitz continuity of F, G, $\hat{f}$ and $\hat{g}$ and the existence and local Lipschitz continuity of φ. In particular, the given functions do not necessarily need to be differentiable or convex in any variable and coupling constraints in the upper level are allowed.

The remainder of the paper is structured as follows. In the next section, we recall some required basic terminology. In Section 3, we discuss the concept of approximate KKT points in a Lipschitzian setting which is the basis for the introduction of the new problem qualification subMFC. The main result in this section shows that any feasible point of (P) is a KKT point provided that it satisfies subMFC. In Section 4, we focus on the comparison of subMFC to existing CQs and the problem qualification calmness. Afterwards, the application to bilevel optimization problems is presented in Section 5. Final remarks are given in the last section of this paper.

2. Preliminaries

Throughout, we use $\|.\|$ to indicate the Euclidean norm. The closed ball centered at $x \in \mathbb{R}^n$ with radius $\epsilon > 0$ is given by $B(x, \epsilon) := \{y \in \mathbb{R}^n \mid \|y - x\| \leq \epsilon\}$. For $x = (x_1, \ldots, x_n)^\top \in \mathbb{R}^n$, we define $x_+ := (\max\{0, x_1\}, \ldots, \max\{0, x_n\})^\top$, whereas $d_C(x)$ stands for the Euclidean distance of x to a set $C \subseteq \mathbb{R}^n$. To shorten notation, from now on a tuple $(x_1, \ldots, x_n)$ of real numbers $x_1, \ldots, x_n \in \mathbb{R}$ will be identified with the vector $(x_1, \ldots, x_n)^\top \in \mathbb{R}^n$. Whenever a function $\phi : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\bar{x} \in \mathbb{R}^n$, we denote the gradient of ϕ at $\bar{x}$ by $\nabla \phi(\bar{x})$. For a function $\phi : \mathbb{R}^n \to \mathbb{R}$ that is Lipschitz continuous near $\bar{x} \in \mathbb{R}^n$, we use the well-known generalization of the gradient by Clarke [13]. More in detail, with the *generalized directional derivative* of ϕ at $\bar{x}$ in direction $d \in \mathbb{R}^n$ given by

$$\phi^\circ(\bar{x}; d) := \limsup_{x \to \bar{x}, h \downarrow 0} \frac{\phi(x + hd) - \phi(x)}{h},$$

the *Clarke subdifferential* of ϕ at $\bar{x}$ is the set

$$\partial \phi(\bar{x}) := \left\{ s \in \mathbb{R}^n \mid s^\top d \leq \phi^\circ(\bar{x}; d) \ \forall d \in \mathbb{R}^n \right\}.$$

If ϕ is continuously differentiable, the Clarke subdifferential reduces to the gradient, i.e., $\partial \phi(\bar{x}) = \{\nabla \phi(\bar{x})\}$. The following proposition provides useful properties of the Clarke subdifferential [13, Proposition 2.1.2, Proposition 2.1.5].

Proposition 2.1. *Let $\phi : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz continuous and $\bar{x} \in \mathbb{R}^n$.*

(a) *The set $\partial \phi(\bar{x})$ is nonempty, convex and compact.*
(b) *If the sequence $\{x^k\}_{k=1}^\infty$ converges to $\bar{x}$ and $s^k \in \partial \phi(x^k)$ for all $k \in \mathbb{N}$, then all accumulation points of the sequence $\{s^k\}_{k=1}^\infty$ belong to $\partial \phi(\bar{x})$.*
(c) *The map $x \mapsto \partial \phi(x)$ is upper semicontinuous at $\bar{x}$, i.e., for all $\epsilon > 0$ there exists $\delta > 0$ such that $\partial \phi(x) \subseteq \partial \phi(\bar{x}) + B(x, \epsilon)$ holds for all $x \in B(\bar{x}, \delta)$.*

Other generalizations of the gradient include the regular and the limiting subdifferential [35, 40]. In detail, let $\phi : \mathbb{R}^n \to \mathbb{R} \cup \{-\infty, \infty\}$ be a lower semicontinuous function that is finite at $\bar{x} \in \mathbb{R}^n$. Then, the *regular/Fréchet subdifferential* of ϕ at $\bar{x}$ is the set

$$\widehat{\partial} \phi(\bar{x}) := \left\{ s \in \mathbb{R}^n \mid \liminf_{x \to \bar{x}, x \neq \bar{x}} \frac{\phi(x) - \phi(\bar{x}) - \langle s, x - \bar{x} \rangle}{\|x - \bar{x}\|} \geq 0 \right\}$$

and the *limiting/Mordukhovich/basic subdifferential* of ϕ at $\bar{x}$ is the set

$$\partial^L \phi(\bar{x}) := \{ \bar{s} \in \mathbb{R}^n \mid \exists \{(x^k, s^k)\}_{k=1}^\infty \subset \mathbb{R}^n \times \mathbb{R}^m : (x^k, s^k) \to (\bar{x}, \bar{s}), \phi(x^k) \to \phi(\bar{x}),$$
$$s^k \in \widehat{\partial} \phi(x^k) \ \forall k \in \mathbb{N}\}.$$

Note that the inclusion $\widehat{\partial} \phi(\bar{x}) \subseteq \partial^L \phi(\bar{x}) \subseteq \partial \phi(\bar{x})$ holds. For a point $\bar{x} \in \mathbb{R}^n$ that is feasible for (P), we define the index set of active constraints as

$$I(\bar{x}) := \{ j \in \{1, \ldots, q\} \mid g_j(\bar{x}) = 0 \}.$$

The nonsmooth extension of the classical MFCQ demands the positive linear independence of the subgradients of all constraints that are active in $\bar{x}$. Thus, its formal definition can be stated as follows.

Definition 2.2. *MFCQ* is said to be satisfied at a point $\bar{x}$ feasible for (P) if $I(\bar{x}) = \emptyset$ or if, for all $v \geq 0$, $v \neq 0$, it holds that

$$0 \notin \sum_{j \in I(\bar{x})} v_j \partial g_j(\bar{x}).$$

3. The Subset Mangasarian-Fromovitz Condition

In this section, we first introduce a suitable sequential optimality condition for the nonsmooth problem (P), which, later on, will constitute a component of subMFC. To this end, let us recall the definitions of Fritz John (FJ) points and KKT points [13, Theorem 6.1.1].

Definition 3.1. A point $\bar{x} \in \mathbb{R}^n$ is called *FJ point* of (P) if there exist multipliers $a \geq 0$, $u \in \mathbb{R}^q$, $(a, u) \neq (0, 0)$ such that

$$0 \in a \partial f(\bar{x}) + \sum_{j=1}^{q} u_j \partial g_j(\bar{x}),$$

$$u \geq 0, \quad g(\bar{x}) \leq 0, \quad u^\top g(\bar{x}) = 0.$$

Definition 3.2. A point $\bar{x} \in \mathbb{R}^n$ is called *KKT point* of (P) if there exists a multiplier $u \in \mathbb{R}^q$ such that

$$0 \in \partial f(\bar{x}) + \sum_{j=1}^{q} u_j \partial g_j(\bar{x}),$$

$$u \geq 0, \quad g(\bar{x}) \leq 0, \quad u^\top g(\bar{x}) = 0.$$

It is known that a local minimizer $\bar{x}$ of problem (P) fulfills the KKT conditions if some CQ is satisfied at $\bar{x}$. But, on the one hand, CQs are often violated, making it challenging to figure out whether the KKT conditions hold at $\bar{x}$. Hence, we instead want to utilize optimality conditions that hold without further requirements. One could therefore be tempted to work with the FJ conditions, as they hold without further assumptions. But, in this case, the multiplier a belonging to the objective function is allowed to vanish, such that the function being minimized is possibly not involved. On the other hand, even if we can expect the KKT conditions to hold, it is generally not possible to compute a KKT point of problem (P) exactly. Hence, numerical algorithms for the treatment of (P) typically incorporate termination criteria that test whether the current iterate is somehow close to a KKT point. Specifically, it is evaluated whether the KKT conditions given in Definition 3.2 are approximately satisfied at the current iterate. This leads to the definition of approximate KKT points, which are well-suited to overcome all the aforementioned drawbacks.

Definition 3.3. A point $\bar{x} \in \mathbb{R}^n$ is called *approximate KKT (AKKT) point* of (P) if there exists a sequence $\{(x^k, u^k, \delta^k, \epsilon^k)\}_{k=1}^{\infty} \subset \mathbb{R}^n \times \mathbb{R}^q \times \mathbb{R}^q \times \mathbb{R}^n$ such that we have

$$\epsilon^k \in \partial f(x^k) + \sum_{j=1}^{q} u_j^k \partial g_j(x^k), \tag{1}$$

$$u^k \geq 0, \quad g(x^k) \leq \delta^k, \quad (u^k)^\top (g(x^k) - \delta^k) = 0 \tag{2}$$

for all $k \in \mathbb{N}$ as well as the convergence

$$(x^k, \delta^k, \epsilon^k) \to (\bar{x}, 0, 0) \quad \text{as } k \to \infty. \tag{3}$$

5

Obviously, every KKT point is also an AKKT point. Definition 3.3 can be seen as the nonsmooth counterpart of the definition in [36]. Note that different definitions of an AKKT point are used in the literature. For example, it can be verified that the definition above is equivalent to the one used in [31]. If we additionally require $\delta^k \geq 0$, the definition obtained in this case would be equivalent to the one used in [11] for the smooth case.

Next, we are going to prove that being an AKKT point is a necessary optimality condition that holds without further assumptions. Related results for nonsmooth settings can be found in [29, 31], where the limiting subdifferential $\partial^L(\cdot)$ is used as a generalization of the gradient for nonsmooth functions. To maintain the paper self-contained, we now state an explicit proof that is tailored to our definition of an AKKT point and that is based on the Clarke subdifferential. Similar proof techniques for related results can be found in [11, 31].

Theorem 3.4. *If $\bar{x}$ is a local minimizer of* (P), *then $\bar{x}$ is an AKKT point of* (P).

Proof. Recall that D denotes the feasible set of problem (P). As f and g are locally Lipschitz continuous on $\mathbb{R}^n$ and $\bar{x}$ is a local minimizer of (P), there exists $\epsilon > 0$ such that f and g are locally Lipschitz continuous on $B(\bar{x}, \epsilon)$ and $f(x) \geq f(\bar{x})$ holds for all $x \in D \cap B(\bar{x}, \epsilon)$. For $k \in \mathbb{N}$, let us consider the problem

$$\min_x f(x) + k \left\| g(x)_+ \right\|^2 + \|x - \bar{x}\|^2 \quad \text{s.t.} \quad x \in B(\bar{x}, \epsilon). \tag{4}$$

Because a locally Lipschitz continuous objective function is to be minimized on a compact domain, the problem has a solution $x^k \in B(\bar{x}, \epsilon)$ for all $k \in \mathbb{N}$ and the sequence $\{x^k\}_{k=1}^\infty$ stays bounded. Therefore, without loss of generality, we may assume that $x^k \to x^\star$ holds for some $x^\star \in B(\bar{x}, \epsilon)$ as $k \to \infty$. By the definition of x^k and the feasibility of $\bar{x}$, it follows that

$$f(x^k) + k\|g(x^k)_+\|^2 + \|x^k - \bar{x}\|^2 \leq f(\bar{x}) + k \left\| g(\bar{x})_+ \right\|^2 + \|\bar{x} - \bar{x}\|^2 = f(\bar{x}) \tag{5}$$

is satisfied for all $k \in \mathbb{N}$. As f is locally Lipschitz continuous on $B(\bar{x}, \epsilon)$, $\{f(x^k)\}_{k=1}^\infty$ is a bounded sequence. Thus, dividing by k and passing to the limit in (5) yields $\|g(x^\star)_+\|^2 = 0$, which is equivalent to $x^\star \in D$. From (5), it follows in particular that

$$f(x^k) + \|x^k - \bar{x}\|^2 \leq f(\bar{x})$$

holds for all $k \in \mathbb{N}$. Using the latter, we obtain

$$f(x^\star) \leq f(x^\star) + \|x^\star - \bar{x}\|^2 = \lim_{k \to \infty} \left(f(x^k) + \|x^k - \bar{x}\|^2 \right) \leq f(\bar{x}) \leq f(x^\star),$$

and hence, $x^\star = \bar{x}$. Thus, $x^k \to \bar{x}$ as $k \to \infty$ implying that x^k lies in the interior of $B(\bar{x}, \epsilon)$ for all $k \in \mathbb{N}$ large enough. Since x^k is a solution to (4), we obtain, using $f_1(x) := \|x - \bar{x}\|^2$ for all $x \in \mathbb{R}^n$, by [13, Proposition 2.3.2] that

$$0 \in \partial(f + k \left\| g_+ \right\|^2 + f_1)(x^k).$$

With the sum rule [13, Proposition 2.3.3] and the chain rule [13, Theorem 2.3.9] for the Clarke subgradient, we get

$$2(\bar{x} - x^k) \in \partial f(x^k) + \sum_{j: g_j(x^k) > 0} 2k g_j(x^k) \partial g_j(x^k).$$

By setting

$$\epsilon^k := 2(\bar{x} - x^k), u_j^k := \begin{cases} 2kg_j(x^k), & \text{if } g_j(x^k) > 0, \\ 0, & \text{if } g_j(x^k) \leq 0 \end{cases} \text{ and } \delta_j^k := \begin{cases} g_j(x^k), & \text{if } g_j(x^k) > 0, \\ \frac{1}{k}, & \text{if } g_j(x^k) \leq 0, \end{cases}$$

it can be checked that the conditions (1)–(3) hold for the sequence $\{(x^k, u^k, \delta^k, \epsilon^k)\}_{k=1}^{\infty}$, and hence, $\bar{x}$ is an AKKT point. □

Thus, as it is possible to show that every local minimizer of problem (P) is the limit of a sequence of points which satisfy the AKKT conditions when the tolerances approach zero, the procedure of utilizing AKKT conditions for termination criteria in numerical algorithms is theoretically justified.

Remark 1. The proof of Theorem 3.4 shows in particular that for a local minimizer $\bar{x}$ of (P), it is always possible to find a sequence $\{(x^k, u^k, \delta^k, \epsilon^k)\}_{k=1}^{\infty}$ with $\delta_j^k > 0$ for all $j = 1, \ldots, q$ and $k \in \mathbb{N}$ for which Definition 3.3 is fulfilled.

Next, we want to proceed with the introduction of subMFC. To this end, for any $x \in \mathbb{R}^n$ that fulfills $g(x) \leq \delta$, we define the index set

$$I(x, \delta) := \{j \in \{1, \ldots, q\} \mid g_j(x) = \delta_j\}.$$

Remark 2. Whenever an AKKT point with a corresponding sequence $\{(x^k, u^k, \delta^k, \epsilon^k)\}_{k=1}^{\infty}$ is considered in the following, we may assume without loss of generality that the associated set of active constraints $I(x^k, \delta^k)$ is constant for all $k \in \mathbb{N}$, which is reasonable as it is always possible to consider a suitable infinite subsequence.

Definition 3.5. Let $\bar{x}$ be feasible for (P). We say that the *Subset Mangasarian-Fromovitz Condition (subMFC)* holds at $\bar{x}$ if there exists $I \subseteq I(\bar{x})$ such that the following two conditions are satisfied:

(i) If $I \neq \emptyset$, it holds for all $v \geq 0, v \neq 0$ that

$$0 \notin \sum_{j \in I} v_j \partial g_j(\bar{x}).$$

(ii) There exists a sequence $\{(x^k, u^k, \delta^k, \epsilon^k)\}_{k=1}^{\infty}$ that fulfills system (1)–(3) with

$$I = I(x^k, \delta^k).$$

By Theorem 3.4, for the special case of $\bar{x}$ being a local minimizer of (P) there always exists at least one sequence that is a suitable candidate for part (ii) in the definition above, as part (ii) means that the considered point $\bar{x}$ is an AKKT point. Because of $I \subseteq I(\bar{x})$, part (i) of the definition can be interpreted as a weakened MFCQ at $\bar{x}$, since we demand the positive linear independence of subgradients of active constraints only for some *subset* of all constraints that are active in $\bar{x}$. However, the positive linear independence of these fewer subgradients already suffices to prove that the AKKT point satisfying (ii) is already a KKT point, as the following theorem shows.

Theorem 3.6. *If $\bar{x}$ is feasible for* (P) *and subMFC holds at $\bar{x}$, then $\bar{x}$ is a KKT point of* (P).

Proof. By Definitions 3.3 and 3.5, it follows that there is a sequence $\{(x^k, u^k, \delta^k, \epsilon^k)\}_{k=1}^{\infty}$ such that (2), $(x^k, \delta^k, \epsilon^k) \to (\bar{x}, 0, 0)$ for $k \to \infty$ and

$$\epsilon^k \in \partial f(x^k) + \sum_{j \in I} u_j^k \partial g_j(x^k) \qquad (6)$$

hold with $I = I(x^k, \delta^k)$, taking into account Remark 2. If $I = \emptyset$, Proposition 2.1(b) and the local Lipschitz continuity of f imply $0 \in \partial f(\bar{x})$ when taking the limit $x^k \to \bar{x}$ in (6). Further, the feasibility of $\bar{x}$ follows from the continuity of g such that we immediately obtain that $\bar{x}$ is a KKT point with multiplier $u := 0$.

For the case $I \neq \emptyset$, we first show that the sequence $\{u^k\}_{k=1}^{\infty}$ is bounded. To this end, let us assume the contrary. Then, without loss of generality, $\|u^k\| \to \infty$ as $k \to \infty$. Now, consider the bounded sequence $\{\bar{u}^k\}_{k=1}^{\infty}$ given by

$$\bar{u}^k := \frac{u^k}{\|u^k\|}.$$

Due to the boundedness, we may assume without loss of generality that $\bar{u}^k \to \bar{u}$ with $\|\bar{u}\| = 1$ and $\bar{u} \geq 0$. Since f is locally Lipschitz continuous and $x^k \to \bar{x}$, we have by parts (a) and (c) of Proposition 2.1 that any sequence $\{s^k\}_{k=1}^{\infty}$ with $s^k \in \partial f(x^k)$ is bounded, which implies

$$\frac{s^k}{\|u^k\|} \to 0 \quad \text{as } k \to \infty.$$

Dividing by $\|u^k\|$ and passing to the limit in (6) therefore yields

$$0 \in \sum_{j \in I} \bar{u}_j \partial g_j(\bar{x}),$$

where we used $\epsilon^k \to 0$ and Proposition 2.1(b) as g is locally Lipschitz continuous. This contradicts subMFC, and hence, the sequence $\{u^k\}_{k=1}^{\infty}$ has to be bounded. Thus, without loss of generality, $u^k \to \tilde{u}$. In turn, as $k \to \infty$, it follows from (2), (6), Proposition 2.1, $\delta^k \to 0$ and the fact that $I \subseteq I(\bar{x})$ by the continuity of g, that $\bar{x}$ is a KKT point of (P) with multiplier $\tilde{u}$. $\qquad \square$

The subsequent example shows that the implication in Theorem 3.6 cannot be inverted, proving that subMFC is not simply a reformulation of the KKT conditions. Moreover, the example reveals that subMFC is not a necessary optimality condition. Rather, although part (ii) of subMFC is a sequential necessary optimality condition that also holds at any KKT point, subMFC as a whole is not. For this reason, it also does not belong to the class of sequential optimality conditions in the classical sense [2, 5, 8, 11]. Thus, the classification of subMFC as problem qualification is justified.

Example 3.7. We consider the Lipschitz continuous problem

$$\begin{aligned}
\min \, 3x_1 - 2x_2 \quad \text{s.t.} \quad & g_1(x_1, x_2) := |x_1| - x_2 \leq 0, \\
& g_2(x_1, x_2) := -|x_1| + x_2 \leq 0, \\
& g_3(x_1, x_2) := x_1^2 + (x_2 + 1)^2 - 1 \leq 0, \\
& g_4(x_1, x_2) := -x_1^2 - (x_2 + 1)^2 + 1 \leq 0
\end{aligned}$$

with its only feasible point $\bar{x} = (\bar{x}_1, \bar{x}_2) = (0, 0)$. First, we see that $\bar{x}$ is a KKT point (and thus also an AKKT point), as

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \begin{pmatrix} 3 \\ -2 \end{pmatrix} + u_1 \begin{pmatrix} [-1, 1] \\ -1 \end{pmatrix} + u_2 \begin{pmatrix} [-1, 1] \\ 1 \end{pmatrix} + u_3 \begin{pmatrix} 0 \\ 2 \end{pmatrix} + u_4 \begin{pmatrix} 0 \\ -2 \end{pmatrix}$$

can be fulfilled by choosing $u = (0, 3, 0, \frac{1}{2})$ and the subgradient $(-1, 1) \in \partial g_2(\bar{x})$. To show that subMFC does not hold at $\bar{x}$, let us pick a sequence $\{(x^k, u^k, \delta^k, \epsilon^k)\}_{k=1}^{\infty}$ satisfying (ii) of subMFC. Then, $(x^k, \delta^k, \epsilon^k) \to (\bar{x}, 0, 0)$ holds for $k \to \infty$ and for all $k \in \mathbb{N}$ we have

$$\begin{pmatrix} \epsilon_1^k \\ \epsilon_2^k \end{pmatrix} \in \begin{pmatrix} 3 \\ -2 \end{pmatrix} + u_1^k \begin{pmatrix} \partial(|x_1^k|) \\ -1 \end{pmatrix} + u_2^k \begin{pmatrix} -\partial(|x_1^k|) \\ 1 \end{pmatrix} + u_3^k \begin{pmatrix} 2x_1^k \\ 2(x_2^k + 1) \end{pmatrix} + u_4^k \begin{pmatrix} -2x_1^k \\ -2(x_2^k + 1) \end{pmatrix}, \quad (7)$$
$$u_i^k \geq 0, \quad g_i(x_1^k, x_2^k) \leq \delta_i^k, \quad u_i^k(g_i(x_1^k, x_2^k) - \delta_i^k) = 0, \quad \forall i = 1, \ldots, 4,$$

where we use $\partial(|x_1^k|)$ as an abbreviation for $\partial|\cdot|(x_1^k)$, i.e., for the subdifferential of the function $x \mapsto |x|$ at x_1^k. We now want to find out for which sets $I \subseteq I(\bar{x}) = \{1, 2, 3, 4\}$ part (i) of subMFC can be satisfied. As we have

$$\partial g_1(\bar{x}) = \begin{pmatrix} [-1, 1] \\ -1 \end{pmatrix}, \partial g_2(\bar{x}) = \begin{pmatrix} [-1, 1] \\ 1 \end{pmatrix}, \nabla g_3(\bar{x}) = \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \nabla g_4(\bar{x}) = \begin{pmatrix} 0 \\ -2 \end{pmatrix},$$

part (i) of subMFC clearly only holds if

$$\text{(a) } I(x^k, \delta^k) \subseteq \{1, 4\} \quad \text{or} \quad \text{(b) } I(x^k, \delta^k) \subseteq \{2, 3\}$$

is valid. This implies for $u^k = (u_1^k, u_2^k, u_3^k, u_4^k)$ that

$$\text{(a) } u_1^k \geq 0, u_2^k = 0, u_3^k = 0, u_4^k \geq 0 \quad \text{or} \quad \text{(b) } u_1^k = 0, u_2^k \geq 0, u_3^k \geq 0, u_4^k = 0.$$

In case (a), we immediately see that the second line of (7) reads as $\epsilon_2^k = -2 - u_1^k - 2(x_2^k + 1)u_4^k$, and hence, $\epsilon_2^k \to 0$ as $k \to \infty$ cannot be satisfied. Therefore, only case (b) has to be considered in the following, meaning that (7) reads as

$$\begin{pmatrix} \epsilon_1^k \\ \epsilon_2^k \end{pmatrix} \in \begin{pmatrix} 3 \\ -2 \end{pmatrix} + u_2^k \begin{pmatrix} -\partial(|x_1^k|) \\ 1 \end{pmatrix} + u_3^k \begin{pmatrix} 2x_1^k \\ 2(x_2^k + 1) \end{pmatrix}. \quad (8)$$

As $x^k \to \bar{x} = (0, 0)$ for $k \to \infty$, we may assume without loss of generality (by passing to a subsequence if necessary) that one of the three cases $x_1^k > 0$, $x_1^k = 0$ or $x_1^k < 0$ holds for all $k \in \mathbb{N}$.

- If $x_1^k > 0$, system (8) reads as

$$\begin{pmatrix} \epsilon_1^k \\ \epsilon_2^k \end{pmatrix} = \begin{pmatrix} 3 \\ -2 \end{pmatrix} + u_2^k \begin{pmatrix} -1 \\ 1 \end{pmatrix} + u_3^k \begin{pmatrix} 2x_1^k \\ 2(x_2^k + 1) \end{pmatrix},$$

leading to

$$\epsilon_2^k = -2 + u_2^k + 2(x_2^k + 1)u_3^k = -2 + 3 - \epsilon_1^k + 2x_1^k u_3^k + 2(x_2^k + 1)u_3^k.$$

The right-hand side is clearly bounded away from 0 for large $k \in \mathbb{N}$ and hence, $\epsilon_2^k \to 0$ as $k \to \infty$ cannot be satisfied.

- If $x_1^k = 0$, equation (8) amounts to

$$\begin{pmatrix} \epsilon_1^k \\ \epsilon_2^k \end{pmatrix} \in \begin{pmatrix} 3 \\ -2 \end{pmatrix} + u_2^k \begin{pmatrix} [-1,1] \\ 1 \end{pmatrix} + u_3^k \begin{pmatrix} 0 \\ 2(x_2^k+1) \end{pmatrix}, \tag{9}$$

which yields $\epsilon_1^k = 3 + u_2^k \cdot s$ for some $s \in [-1,1]$. To guarantee $\epsilon_1^k \to 0$ as $k \to \infty$, only $s \in [-1,0)$ is possible. Thus, we obtain $u_2^k = (\epsilon_1^k - 3)s^{-1}$ for some $s \in [-1,0)$. Plugging the latter into the second line of (9) gives

$$\epsilon_2^k = -2 + u_2^k + 2(x_2^k+1)u_3^k = -2 + (\epsilon_1^k-3)s^{-1} + 2(x_2^k+1)u_3^k \geq -2 + 3 - \epsilon_1^k + 2(x_2^k+1)u_3^k.$$

This again is bounded away from 0 for large $k \in \mathbb{N}$, such that $\epsilon_2^k \to 0$ as $k \to \infty$ cannot hold.

- In the case $x_1^k < 0$, equation (8) reads as

$$\begin{pmatrix} \epsilon_1^k \\ \epsilon_2^k \end{pmatrix} = \begin{pmatrix} 3 \\ -2 \end{pmatrix} + u_2^k \begin{pmatrix} 1 \\ 1 \end{pmatrix} + u_3^k \begin{pmatrix} 2x_1^k \\ 2(x_2^k+1) \end{pmatrix},$$

which yields $\epsilon_2^k = -2 + u_2^k + 2(x_2^k+1)u_3^k$ and thus

$$u_3^k = \frac{2 + \epsilon_2^k - u_2^k}{2(x_2^k+1)}, \quad \epsilon_1^k = 3 + u_2^k + x_1^k \frac{2+\epsilon_2^k-u_2^k}{x_2^k+1} = 3 + \left(1 - \frac{x_1^k}{x_2^k+1}\right)u_2^k + \frac{x_1^k(2+\epsilon_2^k)}{x_2^k+1}.$$

Clearly, it is again not possible to satisfy $\epsilon_1^k \to 0$ as $k \to \infty$.

Thus, it is impossible to satisfy both (i) and (ii) of subMFC at the same time. Altogether, this means that subMFC cannot hold at the KKT point $\bar{x}$.

Remark 3. As Theorem 3.6 holds for any $\bar{x}$ that is feasible for (P), we can state analogous results for $\bar{x}$ being a local minimizer or for $\bar{x}$ being an AKKT point of (P). For these special cases, the theorem reveals that subMFC can be used as a condition to guarantee that a local minimizer (resp. an AKKT point) fulfills the KKT conditions. This essentially means that subMFC still shares the defining property of a CQ (resp. of a strict CQ), even though it is not a CQ itself.

Remark 4. Of course, we could also state Definitions 3.3 and 3.5 with respect to other sub-differentials than Clarke's subdifferential, such as the regular subdifferential $\widehat{\partial}(\cdot)$ or the limiting subdifferential $\partial^L(\cdot)$. As these subdifferentials are generally smaller than Clarke's sub-differential, the definition of an AKKT point would become more restrictive. For subMFC, the positive linear independence required for part (i) would generally be easier to fulfill, whereas part (ii) would become harder to fulfill. Hence, it is not straightforward to de-rive a general rule whether subMFC becomes weaker or stronger when using a different subdifferential.

To end this section, we want to illustrate how subMFC can be applied. Simultaneously, this part also clarifies how subMFC can be used to treat problems with equality constraints. For this purpose, we first take a look at a standard equality constrained example in which MFCQ fails to demonstrate the benefit of using subMFC.

Example 3.8. Consider an optimization problem with the feasible set given by the equality constraint $x^2 = 0$, which can equivalently be written as feasible set of an instance of optimization problem (P) by splitting the equation into two inequalities:

$$\min x^2 \quad \text{s.t.} \quad g_1(x) := x^2 \le 0, \ g_2(x) := -x^2 \le 0.$$

In the only feasible point and global minimizer $\bar{x} = 0$, we obtain $I(\bar{x}) = \{1, 2\}$. Clearly, MFCQ does not hold at $\bar{x}$. Concerning subMFC, we first take a look at the AKKT conditions in $\bar{x}$ for the above problem, which require $(x^k, \delta^k, \epsilon^k) \to (\bar{x}, 0, 0)$ for $k \to \infty$ as well as

$$\epsilon^k = 2x^k + u_1^k(2x^k) + u_2^k(-2x^k),$$
$$u_1^k \ge 0, \quad (x^k)^2 \le \delta_1^k, \quad u_1^k((x^k)^2 - \delta_1^k) = 0,$$
$$u_2^k \ge 0, \quad -(x^k)^2 \le \delta_2^k, \quad u_2^k(-(x^k)^2 - \delta_2^k) = 0$$

for all $k \in \mathbb{N}$. To satisfy subMFC at $\bar{x}$, we need to find a sequence $\{(x^k, u^k, \delta^k, \epsilon^k)\}_{k=1}^{\infty}$ that fulfills the above conditions such that either $I^g(x^k, \delta^k) = \emptyset$ or

$$0 \ne \sum_{j \in I^g(x^k, \delta^k)} v_j \nabla g_j(\bar{x})$$

holds for all $v \ge 0$, $v \ne 0$. Due to $\nabla g_1(\bar{x}) = 0$ and $\nabla g_2(\bar{x}) = 0$, this implies that we need $I^g(x^k, \delta^k) = \emptyset$ for subMFC to hold at $\bar{x}$. Indeed, a sequence fulfilling both the AKKT conditions and $I^g(x^k, \delta^k) = \emptyset$ can be found, e.g.

$$\{(x^k, u^k, \delta^k, \epsilon^k)\}_{k=1}^{\infty} := \left\{0, (0, 0), \left(\frac{1}{k}, \frac{1}{k}\right), 0\right\}_{k=1}^{\infty}.$$

Thus, the example shows that we can exclude unwanted gradients from the index set $I^g(x^k, \delta^k)$ by setting δ^k accordingly. The same idea can also be applied if the point of interest is not the unrestricted minimizer of the objective function, i.e., when $I^g(x^k, \delta^k) \ne \emptyset$ is required, as the following example shows.

Example 3.9. Consider an optimization problem with the objective function $f(x_1, x_2) := x_1 - 2x_2$ and the feasible set given by the equality constraints $h_1(x) = h_1(x_1, x_2) := |x_1| - x_2 = 0$ and $h_2(x) = h_2(x_1, x_2) := x_1^2 + (x_2 + 1)^2 - 1 = 0$, for which the only feasible point is $\bar{x} = (\bar{x}_1, \bar{x}_2) = (0, 0)$. We obtain the (sub)gradients

$$\partial h_1(x) = \begin{pmatrix} \partial(|x_1|) \\ -1 \end{pmatrix} \text{ with } \partial(|x_1|) = \begin{cases} -1, & \text{for } x_1 < 0, \\ [-1, 1], & \text{for } x_1 = 0, \text{ and } \nabla h_2(x) = \begin{pmatrix} 2x_1 \\ 2(x_2 + 1) \end{pmatrix}, \\ 1, & \text{for } x_1 > 0 \end{cases}$$

amounting to

$$\partial h_1(\bar{x}) = \begin{pmatrix} [-1, 1] \\ -1 \end{pmatrix} \text{ and } \nabla h_2(\bar{x}) = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$$

at $\bar{x}$. Since the (sub)gradients $(0, -1) \in \partial h_1(\bar{x})$ and $\nabla h_2(\bar{x})$ are linearly dependent, LICQ does not hold at $\bar{x}$. In turn, even MFCQ tailored to problems including equality constraints is violated at $\bar{x}$.

The reformulation of the problem into an instance of (P), obtained by splitting the equations into two inequalities, is given by the constraint set in Example 3.7 together with the objective function $f(x_1, x_2) = x_1 - 2x_2$. To verify subMFC at $\bar{x}$, we write down the AKKT conditions at $\bar{x}$, which consist of $(x^k, \delta^k, \epsilon^k) \to (\bar{x}, 0, 0)$ as $k \to \infty$ and

$$\begin{pmatrix} \epsilon_1^k \\ \epsilon_2^k \end{pmatrix} \in \begin{pmatrix} 1 \\ -2 \end{pmatrix} + u_1^k \begin{pmatrix} \partial(|x_1^k|) \\ -1 \end{pmatrix} + u_2^k \begin{pmatrix} -\partial(|x_1^k|) \\ 1 \end{pmatrix} + u_3^k \begin{pmatrix} 2x_1^k \\ 2(x_2^k + 1) \end{pmatrix} + u_4^k \begin{pmatrix} -2x_1^k \\ -2(x_2^k + 1) \end{pmatrix},$$
$$u_i^k \geq 0, \quad g_i(x_1^k, x_2^k) \leq \delta_i^k, \quad u_i^k(g_i(x_1^k, x_2^k) - \delta_i^k) = 0, \quad \forall i = 1, \dots, 4$$

for all $k \in \mathbb{N}$. It can be checked that one sequence to fulfill this system is given by

$$\{(x^k, u^k, \delta^k, \epsilon^k)\}_{k=1}^\infty := \left\{ \left(\left(\frac{1}{k}, 0 \right), \left(0, 1, \frac{1}{2}, 0 \right), \left(\frac{2}{k}, -\frac{1}{k}, \frac{1}{k^2}, 0 \right), \left(\frac{1}{k}, 0 \right) \right) \right\}_{k=1}^\infty.$$

For this sequence we have $I(x^k, \delta^k) = \{2, 3\}$ for all $k \in \mathbb{N}$. Thus, setting $I := \{2, 3\}$, we see that

$$0 \notin \sum_{j \in I} v_j \partial g_j(\bar{x}) = v_2 \begin{pmatrix} -\partial(|\bar{x}_1|) \\ 1 \end{pmatrix} + v_3 \begin{pmatrix} 2\bar{x}_1 \\ 2(\bar{x}_2 + 1) \end{pmatrix} = v_2 \begin{pmatrix} [-1, 1] \\ 1 \end{pmatrix} + v_3 \begin{pmatrix} 0 \\ 2 \end{pmatrix}$$

is satisfied for all $(v_2, v_3) \neq (0, 0)$ with $v_2 \geq 0, v_3 \geq 0$, i.e., subMFC holds at $\bar{x}$.

In summary, the constraints that are active in $\bar{x}$ have linearly dependent (sub)gradients, which leads to a violation of MFCQ, no matter whether the constraints are posed as equality constraints or as split inequality constraints. On the opposite, the split into inequality constraints allows subMFC to choose an appropriate selection of the (sub)gradients. Precisely, we only pick those linearly independent (sub)gradients which are crucial for fulfilling the AKKT conditions, while all others are neglected by choosing the corresponding component of δ^k accordingly. In particular, we have seen that subMFC is able to identify KKT points starting from AKKT points in situations where standard CQs are not readily available. The following section is dedicated to showing that this benefit also occurs in situations where far weaker CQs and problem qualifications are violated.

4. Comparison to constraint qualifications and calmness

In order to show that subMFC is a reasonable and powerful condition, the next aim is to prove in Sections 4.1–4.3 that it is automatically fulfilled for a nontrivial class of optimization problems, as many well-known CQs strictly imply subMFC at local minimizers. Thereafter, we demonstrate in Sections 4.4–4.7 that several weak conditions are not implied by subMFC or known to be independent of subMFC. For a clearer presentation of this part, the corresponding detailed calculations are provided in the appendix. Finally, in Section 4.8, a summary of the investigated dependencies is given.

4.1. Mangasarian-Fromovitz constraint qualification

Recall the definition of MFCQ, Definition 2.2. As indicated before, subMFC can be interpreted as a weakened MFCQ which aligns with the next result.

Theorem 4.1. *Let $\bar{x}$ be a local minimizer of* (P). *Then, MFCQ at $\bar{x}$ strictly implies subMFC at $\bar{x}$.*

Proof. As $\bar{x}$ is an AKKT point by Theorem 3.4, there exists a sequence $\{(x^k, u^k, \delta^k, \epsilon^k)\}_{k=1}^\infty$ that fulfills system (1)–(3). Recalling Remark 2, we can take $I := I(x^k, \delta^k)$ to fulfill part (ii) of subMFC at $\bar{x}$. By the continuity of g and due to $\delta^k \to 0$, it follows that $I \subseteq I(\bar{x})$. Thus, if $I(\bar{x}) = \emptyset$, then $I = \emptyset$ such that part (i) of subMFC is fulfilled automatically. Otherwise, if $I(\bar{x}) \neq \emptyset$, then by MFCQ we have for all $v \geq 0, v \neq 0$ that

$$0 \notin \sum_{j \in I(\bar{x})} v_j \partial g_j(\bar{x}) \overset{I \subseteq I(\bar{x})}{\Longrightarrow} 0 \notin \sum_{j \in I} v_j \partial g_j(\bar{x}).$$

Hence, part (i) of subMFC holds at $\bar{x}$ as well. Finally, it follows from Example 3.8 that this implication is strict. $\qquad\square$

Related to MFCQ, a well-known problem qualification in standard nonlinear programming is the strict Mangasarian-Fromovitz CQ (SMFCQ) which, given a KKT point, ensures uniqueness of the associated Lagrange multipliers [23]. Clearly, as SMFCQ presumes the existence of a Lagrange multiplier, it depends on the objective function and, thus, is not a CQ but a problem qualification. To emphasize this circumstance, it was also named strict Mangasarian-Fromovitz condition (SMFC) later on [42]. Since this problem qualification is weaker than LICQ but stronger than MFCQ [23], it trivially implies subMFC as well.

4.2. Constant positive linear dependence constraint qualification

We can adapt the more general CPLD definition from [43, Definition 1.1] for nonsmooth problems to our setting in (P) as follows.

Definition 4.2. *CPLD* is said to hold at a point $\bar{x}$ feasible for (P) if, for any nonempty subset $J \subseteq I(\bar{x})$ and any choice of subgradients $s_j \in \partial g_j(\bar{x}), j \in J$, such that there exists $v \geq 0, v \neq 0$ with

$$0 = \sum_{j \in J} v_j s_j,$$

then, for all $k \in \mathbb{N}$ sufficiently large and for all sequences $\{x^k\}_{k=1}^\infty$, $\{s^k\}_{k=1}^\infty$ satisfying $x^k \neq \bar{x}$, $x^k \to \bar{x}$ as $k \to \infty$, $s_j^k \in \partial g_j(x^k)$ and $s_j^k \to s_j$ as $k \to \infty$, there exists $v^k \neq 0$ such that

$$0 = \sum_{j \in J} v_j^k s_j^k.$$

It is known that CPLD is a CQ [43] and, by definition, it is straightforward to see that MFCQ implies CPLD. In [43, Proposition 4.1], it is further shown that CPLD is also implied by CRCQ. Both of these implications are strict [7]. To prove that CPLD implies subMFC at local minimizers, we make use of the following auxiliary result [11, Lemma 3.1] that can be viewed as a corollary of Carathéodory's Lemma.

Lemma 4.3. *Assume that $v = \sum_{j=1}^q \lambda_j v_j$ with $v_j \in \mathbb{R}^n$ and $\lambda_j \geq 0$ for all $j = 1, \ldots, q$. Then, there exist $J \subseteq \{1, \ldots, q\}$ and $\bar{\lambda}_j \geq 0$ for all $j \in J$ such that:*

(a) $v = \sum_{j \in J} \bar{\lambda}_j v_j$;
(b) The vectors $\{v_j\}_{j \in J}$ are linearly independent.

The next theorem shows that CPLD implies subMFC at a local minimizer if the subdifferentials of the active constraints are singletons in that local minimizer.

Theorem 4.4. *Let $\bar{x}$ be a local minimizer of* (P). *If $\partial g_j(\bar{x})$ is a singleton for all $j \in I(\bar{x})$, then CPLD at $\bar{x}$ strictly implies subMFC at $\bar{x}$.*

Proof. Since $\bar{x}$ is an AKKT point by Theorem 3.4, there is a sequence $\{(x^k, u^k, \delta^k, \epsilon^k)\}_{k=1}^{\infty}$ that fulfills (1)–(3). This yields in particular

$$\epsilon^k = t^k + \sum_{j \in I(x^k, \delta^k)} u_j^k s_j^k \qquad \text{for some } t^k \in \partial f(x^k), \ s_j^k \in \partial g_j(x^k).$$

By Lemma 4.3 and Remark 2, we may assume that there exist an index set J, an infinite index set $N \subseteq \mathbb{N}$ such that $J \subseteq I(x^k, \delta^k)$ for all $k \in N$, and multipliers $\bar{u}^k \geq 0$ for all $k \in N$ such that:

(a) $\epsilon^k = t^k + \sum_{j \in J} \bar{u}_j^k s_j^k$;
(b) The vectors $\{s_j^k\}_{j \in J}$ are linearly independent.

Suppose now that subMFC does not hold at $\bar{x}$. Then, there exists $v \geq 0, v \neq 0$ such that

$$0 = \sum_{j \in I(x^k, \delta^k)} v_j s_j,$$

where $\partial g_j(\bar{x}) = \{s_j\}$ for all $j \in I(x^k, \delta^k)$ by assumption, as we have $I(x^k, \delta^k) \subseteq I(\bar{x})$ by $\delta^k \to 0$ and the continuity of g. Moreover, by parts (a) and (c) of Proposition 2.1, we find that, for all $j \in I(x^k, \delta^k)$, the sequence $\{s_j^k\}_{k=1}^{\infty}$ is convergent for $k \to \infty$ without loss of generality. Due to Proposition 2.1(b), $x^k \to \bar{x}$ and $\partial g_j(\bar{x}) = \{s_j\}$, this particularly means that $s_j^k \to s_j$ for $k \to \infty$. As CPLD holds at $\bar{x}$, the vectors $\{s_j^k\}_{j \in I(x^k, \delta^k)}$ must therefore be linearly dependent.
If $J = I(x^k, \delta^k)$, this is a contradiction to (b), and hence, subMFC holds at $\bar{x}$. Otherwise, we have $I(x^k, \delta^k) \setminus J \neq \emptyset$. As this set contains the indices of $I(x^k, \delta^k)$ that are not used in representation (a), we set

$$(\widetilde{u}_j^k, \widetilde{\delta}_j^k) := \begin{cases} (0, c\delta_j^k) & \forall j \in I(x^k, \delta^k) \setminus J \text{ with } \delta_j^k > 0, \\ (0, -c\delta_j^k) & \forall j \in I(x^k, \delta^k) \setminus J \text{ with } \delta_j^k < 0, \\ (0, \frac{1}{k}) & \forall j \in I(x^k, \delta^k) \setminus J \text{ with } \delta_j^k = 0, \\ (\bar{u}_j^k, \delta_j^k) & \forall j \in J, \\ (0, \delta_j^k) & \forall j \notin I(x^k, \delta^k) \end{cases}$$

for some $c > 1$. Then, the sequence $\{(x^k, \widetilde{u}^k, \widetilde{\delta}^k, \epsilon^k)\}_{k=1}^{\infty}$ fulfills

$$\epsilon^k \in \partial f(x^k) + \sum_{j=1}^{q} \widetilde{u}_j^k \partial g_j(x^k), \tag{10}$$

$$g(x^k) \leq \widetilde{\delta}^k, \quad \widetilde{u}^k \geq 0, \quad (\widetilde{u}^k)^\top (g(x^k) - \widetilde{\delta}^k) = 0, \tag{11}$$

$$(x^k, \widetilde{\delta}^k, \epsilon^k) \to (\bar{x}, 0, 0) \quad \text{as } k \to \infty. \tag{12}$$

Here, (10) follows from (a). Additionally, (11) follows from $\delta_j^k \leq \widetilde{\delta}_j^k$ and the fact that $\widetilde{u}_j^k > 0$ is possible only for $j \in J \subseteq I(x^k, \delta^k)$ for which we have $g_j(x^k) - \widetilde{\delta}_j^k = g_j(x^k) - \delta_j^k = 0$. For the new perturbation vector $\widetilde{\delta}^k$ defined above, it holds that

$$g_j(x^k) - \widetilde{\delta}_j^k < g_j(x^k) - \delta_j^k = 0 \qquad \forall j \in I(x^k, \delta^k) \setminus J,$$
$$g_j(x^k) - \widetilde{\delta}_j^k = g_j(x^k) - \delta_j^k = 0 \qquad \forall j \in J,$$
$$g_j(x^k) - \widetilde{\delta}_j^k = g_j(x^k) - \delta_j^k < 0 \qquad \forall j \notin I(x^k, \delta^k),$$

and hence, $I(x^k, \widetilde{\delta}^k) = J$. The beginning of the proof can now be repeated for the new sequence $\{(x^k, \widetilde{u}^k, \widetilde{\delta}^k, \epsilon^k)\}_{k=1}^\infty$ as (10)–(12) corresponds to (1)–(3) being fulfilled with $\widetilde{u}^k$ and $\widetilde{\delta}^k$ instead of u^k and δ^k. As a consequence, we obtain that subMFC holds at $\bar{x}$.

Finally, we demonstrate that this implication is strict. To this end, consider the optimization problem from Example 3.8, for which it is already known that subMFC holds at the global minimizer $\bar{x} = 0$. Thus, we now verify that CPLD is violated at $\bar{x}$. According to Definition 4.2, we have to look at any nonempty subset $J \subseteq I(\bar{x}) = \{1, 2\}$. For the choice $J := \{1\}$, we find $\nabla g_1(\bar{x}) = 0$, and hence, there exists $v_1 > 0$ with $0 = v_1 \cdot \nabla g_1(\bar{x})$. The sequences $\{x^k\}_{k=1}^\infty$ and $\{s^k\}_{k=1}^\infty$ given by $x^k := \frac{1}{k}$ and $s^k := \frac{2}{k}$ satisfy $x^k \neq 0$, $x^k \to 0$ as $k \to \infty$, $s^k = \nabla g_1(x^k)$ and $s^k \to s$ as $k \to \infty$. However, there exists no $v_1^k \neq 0$ such that for all $k \in \mathbb{N}$ sufficiently large we have $0 = v_1^k \cdot s^k$. Hence, CPLD does not hold. $\qquad\square$

Note that the assumption that $\partial g_j(\bar{x})$ is a singleton for all $j \in I(\bar{x})$ is required as otherwise it would not be guaranteed that the subgradients required for the AKKT conditions and the subgradients required in the assumed contradiction to subMFC coincide. However, taking a closer look at the proof, we can relax this assumption slightly.

Corollary 4.5. *Let $\bar{x}$ be a local minimizer of* (P). *Let $\{(x^k, u^k, \delta^k, \epsilon^k)\}_{k=1}^\infty$ be a sequence such that Definition 3.3 is fulfilled at $\bar{x}$ with $I(x^k, \delta^k)$ being constant for all $k \in \mathbb{N}$. If $\partial g_j(\bar{x})$ is a singleton for all $j \in I(x^k, \delta^k)$, then CPLD at $\bar{x}$ strictly implies subMFC at $\bar{x}$.*

As $I(x^k, \delta^k) \subseteq I(\bar{x})$ holds due to the continuity of g and $\delta^k \to 0$, Corollary 4.5 generally requires less subdifferentials to be a singleton than Theorem 4.4. Moreover, the additional assumptions concerning the concrete sequence are not restrictive since they can always be fulfilled, see Theorem 3.4 and Remark 2.

4.3. Pseudonormality and quasinormality

Next, we consider the well-known definitions of pseudonormality and quasinormality, introduced in [10] and [21]. They can be extended to the Lipschitz continuous setting in (P) as follows.

Definition 4.6. A point $\bar{x}$ feasible for (P) is said to satisfy

(a) *pseudonormality* if there exists no multiplier $v \geq 0, v \neq 0$ such that:

(a1) $0 \in \sum_{j \in I(\bar{x})} v_j \partial g_j(\bar{x})$;

(a2) There exists a sequence $\{x^k\}_{k=1}^\infty$ with $x^k \to \bar{x}$ as $k \to \infty$ such that

$$\sum_{j \in I(\bar{x})} v_j g_j(x^k) > 0$$

holds for all $k \in \mathbb{N}$.

(b) *quasinormality* if there exists no multiplier $v \geq 0, v \neq 0$ such that:

(b1) $0 \in \sum_{j \in I(\bar{x})} v_j \partial g_j(\bar{x})$;

(b2) There exists a sequence $\{x^k\}_{k=1}^{\infty}$ with $x^k \to \bar{x}$ as $k \to \infty$ such that

$$v_j g_j(x^k) > 0 \quad \forall j \text{ with } v_j > 0$$

holds for all $k \in \mathbb{N}$.

The definitions immediately show that both conditions are weaker than MFCQ. Moreover, they show that pseudonormality implies quasinormality, where the reverse statement is not true by [10, Example 3.1]. If all constraint functions $g_j, j = 1, \ldots, q$, of (P) are concave, then any feasible point of (P) satisfies pseudonormality [45, Proposition 3]. Further, in [7] it is shown that CPLD neither implies nor is implied by pseudonormality. In contrast, quasinormality is implied by CPLD for the smooth case [7, Theorem 3.1]. However, this implication does not hold anymore for the nonsmooth case [39], which is the reason why we considered CPLD separately in Section 4.2. Fortunately, it turns out that even quasinormality implies subMFC.

Theorem 4.7. *Let $\bar{x}$ be a local minimizer of* (P)*. Then, quasinormality at $\bar{x}$ strictly implies subMFC at $\bar{x}$.*

Proof. From Remarks 1 and 2 it follows that there exists a sequence $\{(x^k, u^k, \delta^k, \epsilon^k)\}_{k=1}^{\infty}$ which fulfills system (1)–(3) with $\delta_j^k > 0$ for all $j = 1, \ldots, q$ and $I(x^k, \delta^k)$ being constant for all $k \in \mathbb{N}$. Suppose that $\bar{x}$ satisfies quasinormality, while subMFC does not hold at $\bar{x}$. For $I := I(x^k, \delta^k)$, it follows that (ii) of subMFC is fulfilled at $\bar{x}$. Hence, as we suppose that subMFC is violated at $\bar{x}$, there must be a contradiction to (i) of subMFC for $I = I(x^k, \delta^k)$. Thus, there exists some $u \geq 0, u \neq 0$ with

$$0 \in \sum_{j \in I} u_j \partial g_j(\bar{x}).$$

As $I = I(x^k, \delta^k) \subseteq I(\bar{x})$ by the continuity of g and due to $\delta^k \to 0$, it follows for

$$v_j := \begin{cases} u_j & \forall j \in I, \\ 0 & \forall j \in I(\bar{x}) \setminus I \end{cases}$$

that the inclusion in (b1) of Definition 4.6 holds. Thus, for all j with $v_j > 0$ we have $j \in I = I(x^k, \delta^k)$, and hence,

$$v_j g_j(x^k) = v_j \delta_j^k > 0,$$

due to $\delta_j^k > 0$ for all $j = 1, \ldots, q$. Hence, as (b2) of Definition 4.6 holds as well, quasinormality is violated at $\bar{x}$. By this contradiction, it follows that quasinormality implies subMFC at $\bar{x}$. Finally, Example 4.8 below shows that this implication is strict. $\qquad \square$

In the subsequent example it is shown that subMFC can serve as a condition to identify a KKT point in a situation where pseudonormality and quasinormality fail.

Example 4.8. Consider the Lipschitzian optimization problem

$$\min (x_1 + 2)^2 + (x_2 - 1)^2 \quad \text{s.t.} \quad g_1(x_1, x_2) := x_2^3 - 3x_2 - \widetilde{g}_1(x_1) \le 0,$$

where

$$\widetilde{g}_1(x_1) := \begin{cases} x_1^3 - 3x_1, & \text{for } x_1 \in (-\infty, -2) \cup (1, \infty), \\ -2, & \text{for } x_1 \in [-2, 1]. \end{cases}$$

The function g_1 is not differentiable but locally Lipschitz continuous. Its Clarke subdifferential is given by

$$\partial g_1(x_1, x_2) = \begin{pmatrix} \partial(-\widetilde{g}_1)(x_1) \\ 3x_2^2 - 3 \end{pmatrix}, \quad \partial(-\widetilde{g}_1)(x_1) = \begin{cases} -3x_1^2 + 3, & \text{for } x_1 \in (-\infty, -2) \cup (1, \infty), \\ [-9, 0], & \text{for } x_1 = -2, \\ 0, & \text{for } x_1 \in (-2, 1]. \end{cases}$$

It is straightforward to see that the global solution to the problem is $\bar{x} = (\bar{x}_1, \bar{x}_2) = (-2, 1)$ and that it is a KKT point with multiplier $u = 0$ and $I(\bar{x}) = \{1\}$. We now take a look at the different conditions pseudonormality, quasinormality and subMFC:

(1) First, we consider quasinormality at $\bar{x}$. It follows from $I(\bar{x}) = \{1\}$ and

$$\partial g_1(\bar{x}_1, \bar{x}_2) = \begin{pmatrix} [-9, 0] \\ 0 \end{pmatrix}$$

that part (b1) of Definition 4.6 is fulfilled for any $v_1 > 0$. By taking the sequence $\{(x_1^k, x_2^k)\}_{k=1}^{\infty}$ with $(x_1^k, x_2^k) := \left(-2 + \frac{1}{k}, 1 + \frac{1}{k}\right) \to (\bar{x}_1, \bar{x}_2)$ as $k \to \infty$, we see that

$$v_1 g_1(x_1^k, x_2^k) = v_1((x_2^k)^3 - 3x_2^k + 2) = v_1 \left(\frac{1}{k^3} + \frac{3}{k^2}\right) > 0 \qquad \forall v_1 > 0.$$

Thus, the point $\bar{x}$ does not satisfy quasinormality which implies that pseudonormality can neither be fulfilled.

(2) Next, we show that subMFC is fulfilled at $\bar{x}$. The AKKT conditions in $\bar{x}$ for this problem require $(x^k, \delta^k, \epsilon^k) \to (\bar{x}, 0, 0)$ for $k \to \infty$ and, for all $k \in \mathbb{N}$,

$$\begin{pmatrix} \epsilon_1^k \\ \epsilon_2^k \end{pmatrix} \in \begin{pmatrix} 2(x_1^k + 2) \\ 2(x_2^k - 1) \end{pmatrix} + u^k \begin{pmatrix} \partial(-\widetilde{g}_1)(x_1^k) \\ 3(x_2^k)^2 - 3 \end{pmatrix},$$

$$u^k \ge 0, \quad (x_2^k)^3 - 3x_2^k - \widetilde{g}_1(x_1^k) \le \delta^k, \quad u^k(x_2^3 - 3x_2 - \widetilde{g}_1(x_1^k) - \delta^k) = 0.$$

It can be easily verified that the sequence

$$\{(x^k, u^k, \delta^k, \epsilon^k)\}_{k=1}^{\infty} := \left\{\left(\left(-2 + \frac{1}{k}, 1 + \frac{1}{k}\right), 0, \frac{6}{k^2}, \left(\frac{2}{k}, \frac{2}{k}\right)\right)\right\}_{k=1}^{\infty}$$

fulfills the above AKKT conditions at $\bar{x}$. For this sequence, $I(x^k, \delta^k) = \emptyset$ holds for all $k \in \mathbb{N}$. Thus, we immediately see that subMFC is satisfied at $\bar{x}$ with $I := \emptyset$.

4.4. Local error bound condition

It is known that CQs may imply local error bounds to the feasible region if the constraint functions are sufficiently smooth, see, for instance, [3] for CPLD, [33] for CRCQ and [34] for quasinormality. Some of these results can, under additional assumptions, be extended to the nonsmooth case. For example, if the constraint functions are subdifferentially regular at the point of interest $\bar{x}$, which means that $\widehat{\partial} g_j(\bar{x}) = \partial^L g_j(\bar{x})$ holds for all $j = 1, \ldots, q$, the local error bound condition is implied by CPLD [43, Theorem 3.2] and by quasinormality [45, Theorem 5]. In [20, Corollary 5.3], it was shown that for the latter result the additional assumption on subdifferential regularity of the constraint functions can even be dropped. Thus, as all of the standard CQs considered so far are stronger than subMFC, we want to take a look at the local error bound condition now. To this end, recall that the feasible set of (P) is denoted by D and that $d_D(x)$ stands for the distance of $x \in \mathbb{R}^n$ to the set D.

Definition 4.9. The *local error bound condition* is said to hold at $\bar{x}$ feasible for (P) if there exist constants $c > 0$ and $r > 0$ such that

$$d_D(x) \leq c \, \|g_+(x)\| \quad \forall x \in B(\bar{x}, r).$$

It turns out that subMFC can hold even if the local error bound condition does not. On the other hand, the local error bound condition may be satisfied even if subMFC is violated.

Proposition 4.10. *Let $\bar{x}$ be feasible for* (P)*. Then, subMFC at $\bar{x}$ and the local error bound condition at $\bar{x}$ are independent conditions.*

Proof. See Appendix A.1. $\qquad \square$

4.5. Cone-continuity property

Next, we want to look at the weakest CQ which implies that an AKKT point is a KKT point. This CQ was established in [6] for smooth optimization problems and called CCP. It has been extended to problems with locally Lipschitz continuous functions in [31], where it was renamed as asymptotic regularity.

Definition 4.11. We say that *CCP* holds at a point $\bar{x}$ feasible for (P) if the set-valued mapping $K : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ defined by

$$K(x) = \left\{ \sum_{j \in I(\bar{x})} v_j \partial g_j(x) \mid v_j \geq 0 \right\}$$

is outer semicontinuous at $\bar{x}$, i.e., if

$$\limsup_{x \to \bar{x}} K(x) := \{\bar{z} \in \mathbb{R}^n \mid \exists \{(x^k, z^k)\}_{k=1}^{\infty} \subset \mathbb{R}^{n+n} : (x^k, z^k) \to (\bar{x}, \bar{z}), z^k \in K(x^k) \ \forall k \in \mathbb{N}\}$$
$$\subset K(\bar{x}). \tag{13}$$

In [31, Lemma 3.12], it is shown that CCP is implied by CPLD, and it follows from [6] that this implication is strict. Moreover, in [6], it is shown that CCP neither implies nor is implied by quasinormality. In [6, 11], it is proven that CCP is the weakest possible strict CQ in the smooth case and in [31], a similar claim is made for the nonsmooth case. Due to Theorem 3.6, we obtain that subMFC also possesses the most important property of strict

CQs, which is guaranteeing that an AKKT point is already a KKT point. At the same time, subMFC neither implies nor is implied by CCP, as we will see in the next proposition. This is reasonable as the key point is to recognize that subMFC is not a CQ, as part (ii) of its definition also depends on the given objective function. In contrast, strict CQs have in common that they can be stated without depending on the specific AKKT point sequence. It turns out that dropping this last property enables subMFC to be such a powerful condition: Indeed, in subMFC, only *one* AKKT point sequence has to be considered, and the defining property (i) only needs to be valid for this particular sequence. Due to this link of the AKKT point sequence with an additional property, subMFC can hold even if usual strict CQs fail, while stile maintaining their most important property. The announced proposition illustrates this circumstance.

Proposition 4.12. *Let $\bar{x}$ be feasible for* (P). *Then, subMFC at $\bar{x}$ and CCP at $\bar{x}$ are independent conditions.*

Proof. See Appendix A.2. □

4.6. Abadie constraint qualification and Guignard constraint qualification

Now, we consider the Abadie CQ (ACQ), which is not a strict CQ, but comparatively weak as it is implied by both CCP [6] and quasinormality [9] for the differentiable case. Moreover, ACQ is equivalent to the local error bound condition if the constraint functions g_j, $j = 1, \ldots, q$, are differentiable and convex [25, Theorem 3.5]. As similar results do not exist for the nonsmooth case, we thus investigate the relation of subMFC to ACQ separately in this section. To this end, we use the definition of the nonsmooth ACQ presented in [38]. It is based on the *contingent cone* to the feasible set D of problem (P) at $\bar{x} \in \mathbb{R}^n$, given by

$$T_D(\bar{x}) := \left\{ d \in \mathbb{R}^n \mid \exists \{t_k\}_{k=1}^{\infty}, \{d^k\}_{k=1}^{\infty} : t_k \downarrow 0, d^k \to d \text{ with } \bar{x} + t_k d^k \in D \ \forall k \in \mathbb{N} \right\},$$

and the *cone of locally constrained directions* to D at $\bar{x} \in \mathbb{R}^n$, which is

$$L_D(\bar{x}) := \left\{ d \in \mathbb{R}^n \mid s^\top d \leq 0 \quad \forall s \in \bigcup_{j \in I(\bar{x})} \partial g_j(\bar{x}) \right\}.$$

Definition 4.13. *ACQ is said to hold at a point $\bar{x}$ feasible for* (P) *if*

$$L_D(\bar{x}) \subseteq T_D(\bar{x}).$$

Moreover, we consider the Guignard CQ (GCQ), which is even weaker than ACQ. Similar to the smooth case, it is derived by replacing the contingent cone in the definition of ACQ with a suitable enlargement, as presented in [38].

Definition 4.14. *GCQ is said to hold at a point $\bar{x}$ feasible for* (P) *if*

$$L_D(\bar{x}) \subseteq \mathrm{cl}\,\mathrm{conv}\,T_D(\bar{x}),$$

where $\mathrm{cl}\,\mathrm{conv}\,T_D(\bar{x})$ is the closure of the convex hull of $T_D(\bar{x})$.

Under mild assumptions, it can be shown that these nonsmooth extensions of the smooth ACQ and GCQ are indeed constraint qualifications for (P), see [38, Theorem 2.3] for details. In the next proposition, it is shown that subMFC is independent of both of these weak CQs.

Proposition 4.15. *Let $\bar{x}$ be feasible for* (P). *Then, subMFC at $\bar{x}$ and ACQ/GCQ at $\bar{x}$ are independent conditions.*

Proof. See Appendix A.3. □

4.7. Calmness in the sense of Clarke

Finally, we consider calmness in the sense of Clarke [13, Definition 6.4.1].

Definition 4.16. Let $\bar{x}$ solve (P). Then, (P) is said to be *calm* at $\bar{x}$ if there exist constants $r > 0$ and $\mu > 0$ such that, for all $\alpha \in B(0, r)$ and for all $x \in B(\bar{x}, r)$ with

$$g(x) + \alpha \leq 0,$$

it follows that

$$f(x) - f(\bar{x}) + \mu \, \|\alpha\| \geq 0.$$

Note that calmness can also be defined for $\bar{x}$ being a local minimizer of (P). It is known that calmness is weaker than MFCQ at local minimizers of (P), see [13, Corollary 5 of Theorem 6.5.2]. Like subMFC, calmness is a problem qualification and not a CQ, as it involves the objective function. This allows calmness to be comparatively weak, but even this condition can be violated while subMFC still holds, as the following proposition shows.

Proposition 4.17. *Let $\bar{x}$ be feasible for* (P). *Then, subMFC at $\bar{x}$ does not imply calmness in the sense of Clarke at $\bar{x}$.*

Proof. See Appendix A.4. □

4.8. Summary of the comparisons

In Figure 1, the implications between the conditions considered in this paper are summarized. In particular, we want to emphasize that all implications are strict. All depicted relations can be verified using the results presented throughout Section 4, together with the examples and results listed for completeness in Appendix B.

20

Figure 1. Strict implications between different conditions considered at a local minimizer of the inequality constrained problem (P) are represented by solid arrows. Additional strict implications and one equivalence in case of differentiability of the constraint functions are visualized by dotted arrows. A crossed out arrow symbolizes that the condition at the start of the arrow does not imply the one at its end.

5. Application to bilevel optimization problems

In this section, we tailor subMFC to the lower-level value function reformulation

$$\min_{x,y} \ F(x,y) \quad \text{s.t.} \quad G(x,y) \leq 0, \quad \hat{g}(x,y) \leq 0, \quad \hat{f}(x,y) - \varphi(x) \leq 0 \qquad \text{(LLVFR)}$$

of the optimistic bilevel optimization problem (BP). We want to emphasize that we do not require any additional assumptions besides the local Lipschitz continuity of the bilevel problem data functions F, G, $\hat{f}$ and $\hat{g}$ and the well-definedness and local Lipschitz continuity of the lower-level value function φ, which are rather weak assumptions in bilevel optimization. To guarantee the local Lipschitz continuity of the lower-level value function φ near some x, many different sufficient conditions are known, see, for example, [46, Proposition 1.2], [13, Corollary 1 of Theorem 6.5.2] or [15, Theorem 3.4], and the references mentioned therein.

For a feasible point (x,y) of problem (LLVFR), we denote the index sets of active con-

21

straints by

$$I^G(x,y) \coloneqq \{i \in \{1,\ldots,p\} \mid G_i(x,y) = 0\},$$
$$I^{\hat{g}}(x,y) \coloneqq \{j \in \{1,\ldots,r\} \mid \hat{g}_j(x,y) = 0\},$$
$$I^{\hat{f}-\varphi}(x,y) \coloneqq \{1 \mid \hat{f}(x,y) - \varphi(x) = 0\} = \{1\}.$$

Note that the last equation is valid because, by definition of φ, $\hat{f} - \varphi$ is inherently active in any point that is feasible for (LLVFR). The latter fact also leads to standard CQs failing for every feasible point of (LLVFR), which is, among others, one reason to relax problem (LLVFR) by

$$\min_{x,y} \; F(x,y) \quad \text{s.t.} \quad G(x,y) \leq 0, \quad \hat{g}(x,y) \leq 0, \quad \hat{f}(x,y) - \varphi(x) \leq \gamma \qquad \text{(LLVFR}_\gamma\text{)}$$

with a perturbation $\gamma > 0$ on the lower-level value function constraint [24, 26, 44]. In contrast to (LLVFR), the constraint including the lower-level value function may or may not be active in any point that is feasible for (LLVFR$_\gamma$), and hence, we define

$$I^{\hat{f}-\varphi}(x,y,\gamma) \coloneqq \begin{cases} \{1\}, & \text{if } \hat{f}(x,y) - \varphi(x) - \gamma = 0, \\ \emptyset, & \text{if } \hat{f}(x,y) - \varphi(x) - \gamma < 0. \end{cases}$$

For several subclasses of the problem class (LLVFR$_\gamma$), standard CQs can hold at feasible points for a fixed value $\gamma > 0$ [24, 26, 44]. Thus, our aim is to build a connection between KKT points of (LLVFR$_\gamma$) and meaningful points of (LLVFR), for which we have the following first result.

Proposition 5.1. *Let $\{\gamma_k\}_{k=1}^\infty$ be a sequence of positive reals such that $\gamma_k \downarrow 0$ as $k \to \infty$. Assume that the point (x^k, y^k) is a KKT point of (LLVFR$_{\gamma_k}$) for each $k \in \mathbb{N}$. Then, any accumulation point of $\{(x^k, y^k)\}_{k=1}^\infty$ is feasible for (LLVFR).*

This result can directly be verified using the continuity of the functions G, $\hat{f}$, $\hat{g}$ and φ. Knowing that a sequence of KKT points of (LLVFR$_\gamma$) converges to a feasible point of (LLVFR) for $\gamma \downarrow 0$, we want to investigate next under which conditions more than just feasibility can be transferred to the limit point, i.e., when is this point an FJ point, an AKKT point or a KKT point for (LLVFR)? For the first part, we obtain the following answer which aligns with [47, Proposition 3.2]. To keep the paper self-contained, we also present a short proof.

Proposition 5.2. *Any point that is feasible for (LLVFR) is an FJ point of (LLVFR) with $a = 0$.*

Proof. By definition of φ we have $\hat{f}(x,y) \geq \varphi(x)$ for all $(x,y) \in \{(x,y) \mid \hat{g}(x,y) \leq 0\}$ and consequently, $\hat{f}(x,y) = \varphi(x)$ holds at all points feasible for (LLVFR). Thus, any point $(\bar{x}, \bar{y})$ feasible for (LLVFR) is a global solution to the problem

$$\min_{x,y} \hat{f}(x,y) - \varphi(x) \quad \text{s.t.} \quad G(x,y) \leq 0, \quad \hat{g}(x,y) \leq 0.$$

As φ is Lipschitz-continuous, we can apply the necessary FJ optimality conditions from Definition 3.1 to the above problem and obtain, for any $(\bar{x}, \bar{y})$ feasible for (LLVFR), the

existence of multipliers $w \geq 0$, $u \geq 0$, $v \geq 0$ with $(w, u, v) \neq (0, 0, 0)$ such that

$$0 \in w\partial(\hat{f} - \varphi)(\bar{x}, \bar{y}) + \sum_{i \in I^G(\bar{x}, \bar{y})} u_i \partial G_i(\bar{x}, \bar{y}) + \sum_{j \in I^g(\bar{x}, \bar{y})} v_j \partial \hat{g}_j(\bar{x}, \bar{y}).$$

Consequently, the relation

$$0 \in a\partial F(\bar{x}, \bar{y}) + \sum_{i \in I^G(\bar{x}, \bar{y})} u_i \partial G_i(\bar{x}, \bar{y}) + \sum_{j \in I^g(\bar{x}, \bar{y})} v_j \partial \hat{g}_j(\bar{x}, \bar{y}) + w\partial(\hat{f} - \varphi)(\bar{x}, \bar{y})$$

holds for $a = 0$, which shows that $(\bar{x}, \bar{y})$ is always an FJ point of (LLVFR), but not necessarily a KKT point. $\qquad\square$

Unfortunately, the combination of Proposition 5.1 and Proposition 5.2 shows that, without further assumptions, all accumulation points of sequences of KKT points of (LLVFR$_\gamma$), computed as $\gamma \downarrow 0$, are only FJ points of (LLVFR). As any feasible point shares the same property, we aim to ensure the convergence to a more meaningful point. Thus, naturally the question arises whether being an AKKT point is a more restrictive property than being an FJ point. The next example shows that this is indeed the case.

Example 5.3. Consider the optimistic bilevel optimization problem

$$\min_{x,y} -y \quad \text{s.t.} \quad y \in S(x) := \operatorname*{argmin}_{y}\{xy \mid -y \leq 0, \ y - 1 \leq 0\}.$$

The lower-level value function φ is given by

$$\varphi(x) = \begin{cases} x, & \text{if } x \leq 0, \\ 0, & \text{if } x > 0. \end{cases}$$

This yields the subdifferential

$$\partial(\hat{f} - \varphi)(x, y) = \begin{pmatrix} h(x, y) \\ x \end{pmatrix} \quad \text{with} \quad h(x, y) = \begin{cases} y - 1, & \text{for } x < 0, \\ [y - 1, y], & \text{for } x = 0, \\ y, & \text{for } x > 0 \end{cases}$$

and the lower-level value function reformulation

$$\min_{x,y} -y \quad \text{s.t.} \quad -y \leq 0, \quad y - 1 \leq 0, \quad xy - \varphi(x) \leq 0.$$

Clearly, the point $(\bar{x}, \bar{y}) := \left(0, \frac{1}{2}\right)$ is feasible to the problem. The AKKT conditions (1)–(3) in $(\bar{x}, \bar{y})$ for the reformulation require $(x^k, y^k, \delta^k, \epsilon^k) \to (\bar{x}, \bar{y}, 0, 0)$ as $k \to \infty$ and

$$\epsilon^k \in \begin{pmatrix} 0 \\ -1 \end{pmatrix} + v_1^k \begin{pmatrix} 0 \\ -1 \end{pmatrix} + v_2^k \begin{pmatrix} 0 \\ 1 \end{pmatrix} + w^k \begin{pmatrix} h(x^k, y^k) \\ x^k \end{pmatrix}, \tag{14}$$

$$v_1^k \geq 0, \qquad -y^k \leq \delta_1^k, \qquad v_1^k(-y^k - \delta_1^k) = 0,$$
$$v_2^k \geq 0, \qquad y^k - 1 \leq \delta_2^k, \qquad v_2^k(y^k - 1 - \delta_2^k) = 0,$$
$$w^k \geq 0, \quad x^k y^k - \varphi(x^k) \leq \delta_3^k, \quad w^k(x^k y^k - \varphi(x^k) - \delta_3^k) = 0$$

for all $k \in \mathbb{N}$. Due to $y^k \to \frac{1}{2}$ and $\delta^k \to 0$, we immediately obtain $v_1^k = v_2^k = 0$ for large $k \in \mathbb{N}$. Thus, the second component in (14) reads as $\epsilon_2^k = -1 + w^k x^k$. As $x^k \to 0$ and $\epsilon^k \to 0$ are required, it follows that this equation cannot be fulfilled. Hence, the point $(\bar{x}, \bar{y})$ is feasible (and thus an FJ point) but not an AKKT point.

Moreover, the sequence

$$\{(x^k, y^k, v^k, w^k, \delta^k, \epsilon^k)\}_{k=1}^\infty := \left\{ \left(\frac{1}{k}, 0, (0,0), k, (0,0,0), (0,0) \right) \right\}_{k=1}^\infty$$

can be used to show that the point $(\widetilde{x}, \widetilde{y}) := (0,0)$ is an AKKT point, as it fulfills the above system. However, it can be checked that $(\widetilde{x}, \widetilde{y})$ is not a KKT point, which is important for the next result.

With Example 5.3, we can establish the following strict relations between the different stationarity concepts considered in this paper.

Proposition 5.4. *Let $(\bar{x}, \bar{y})$ be feasible for (LLVFR). Then, the following implications are strict:*

$$(\bar{x}, \bar{y}) \text{ is a KKT point} \Longrightarrow (\bar{x}, \bar{y}) \text{ is an AKKT point} \Longrightarrow (\bar{x}, \bar{y}) \text{ is an FJ point.}$$

Proof. From Definitions 3.2 and 3.3 it follows that, even for the general problem (P), any KKT point $\bar{x}$ with multiplier u is an AKKT point with the corresponding sequence $(x^k, u^k, \delta^k, \epsilon^k) := (\bar{x}, u, 0, 0)$. The second implication holds for (LLVFR) due to Proposition 5.2 and the fact that any AKKT point is in particular feasible, as the constraint functions are continuous. The strictness of the implications then follows from Example 5.3. $\square$

As the following proposition shows, it turns out that no additional assumptions are required to guarantee that an accumulation point of a sequence of KKT points of problem (LLVFR$_\gamma$) for $\gamma \downarrow 0$ fulfills the stronger requirement of being an AKKT point of (LLVFR).

Proposition 5.5. *Let $\{\gamma_k\}_{k=1}^\infty$ be a sequence of positive reals such that $\gamma_k \downarrow 0$ as $k \to \infty$. Assume that the point (x^k, y^k) is a KKT point of (LLVFR$_{\gamma_k}$) for each $k \in \mathbb{N}$. Then, any accumulation point of $\{(x^k, y^k)\}_{k=1}^\infty$ is an AKKT point of (LLVFR).*

Proof. Consider an accumulation point $(\bar{x}, \bar{y})$ of the sequence $\{(x^k, y^k)\}_{k=1}^\infty$. Without loss of generality, we may assume that the whole sequence $\{(x^k, y^k)\}_{k=1}^\infty$ converges to $(\bar{x}, \bar{y})$. As (x^k, y^k) is a KKT point of (LLVFR$_{\gamma_k}$), there are multipliers u^k, v^k and w^k such that

$$0 \in \partial F(x^k, y^k) + \partial G(x^k, y^k) u^k + \partial \hat{g}(x^k, y^k) v^k + w^k \partial (\hat{f} - \varphi)(x^k, y^k),$$

$$u^k \geq 0, \qquad G(x^k, y^k) \leq 0, \qquad (u^k)^\top G(x^k, y^k) = 0,$$

$$v^k \geq 0, \qquad \hat{g}(x^k, y^k) \leq 0, \qquad (v^k)^\top \hat{g}(x^k, y^k) = 0,$$

$$w^k \geq 0, \quad \hat{f}(x^k, y^k) - \varphi(x^k) \leq \gamma_k, \quad w^k(\hat{f}(x^k, y^k) - \varphi(x^k) - \gamma_k) = 0$$

holds for all $k \in \mathbb{N}$. By the assumptions, we further have

$$(x^k, y^k, \gamma_k) \to (\bar{x}, \bar{y}, 0) \quad \text{as } k \to \infty.$$

Consequently, the definition of an AKKT point is fulfilled at $(\bar{x}, \bar{y})$ with the sequence $\{(\widehat{x}^k, \widehat{u}^k, \widehat{\delta}^k, \widehat{\epsilon}^k)\}_{k=1}^\infty$ for $\widehat{x}^k := (x^k, y^k)$, $\widehat{u}^k := (u^k, v^k, w^k)$, $\widehat{\delta}^k := (0, 0, \gamma_k)$ and $\widehat{\epsilon}^k := (0, 0)$. $\square$

Finally, we are interested in finding a suitable condition under which we obtain a KKT point of (LLVFR) when looking at a sequence of KKT points of problem (LLVFR$_\gamma$) for $\gamma \downarrow 0$. As a consequence of Proposition 5.5, we can use the results from Section 3 to establish a connection between KKT points of (LLVFR$_\gamma$) and KKT points of (LLVFR). To this end, we introduce a new problem qualification for the problem class (LLVFR) by tailoring Definition 3.5.

Definition 5.6. Let $(\bar{x}, \bar{y})$ be feasible for (LLVFR). We say that the *LLVFR-subMFC* holds at $(\bar{x}, \bar{y})$ if there exist $I_1 \subseteq I^G(\bar{x}, \bar{y})$, $I_2 \subseteq I^{\hat{g}}(\bar{x}, \bar{y})$ and $I_3 \subseteq I^{\hat{f} - \varphi}(\bar{x}, \bar{y}) = \{1\}$ such that the following two conditions are satisfied:

(i) Either $I_1 = I_2 = I_3 = \emptyset$, or it holds for all $u \geq 0$, $v \geq 0$ and $w \geq 0$ with $(u, v, w) \neq (0, 0, 0)$ that

$$0 \notin \sum_{i \in I_1} u_i \partial G_i(\bar{x}, \bar{y}) + \sum_{j \in I_2} v_j \partial \hat{g}_j(\bar{x}, \bar{y}) + \begin{cases} w \partial(\hat{f} - \varphi)(\bar{x}, \bar{y}), & \text{if } I_3 = \{1\}, \\ 0, & \text{if } I_3 = \emptyset. \end{cases}$$

(ii) There exists a sequence $\{(x^k, y^k, u^k, v^k, w^k, \delta^k, \hat{\delta}^k, \gamma_k, \epsilon^k)\}_{k=1}^\infty$ that fulfills

$$\epsilon^k \in \partial F(x^k, y^k) + \partial G(x^k, y^k) u^k + \partial \hat{g}(x^k, y^k) v^k + w^k \partial(\hat{f} - \varphi)(x^k, y^k), \tag{15}$$

$$u^k \geq 0, \qquad G(x^k, y^k) \leq \delta^k, \qquad (u^k)^\top (G(x^k, y^k) - \delta^k) = 0, \tag{16}$$

$$v^k \geq 0, \qquad \hat{g}(x^k, y^k) \leq \hat{\delta}^k, \qquad (v^k)^\top (\hat{g}(x^k, y^k) - \hat{\delta}^k) = 0, \tag{17}$$

$$w^k \geq 0, \quad \hat{f}(x^k, y^k) - \varphi(x^k) \leq \gamma_k, \quad w^k(\hat{f}(x^k, y^k) - \varphi(x^k) - \gamma_k) = 0 \tag{18}$$

for all $k \in \mathbb{N}$ and

$$(x^k, y^k, \delta^k, \hat{\delta}^k, \gamma_k, \epsilon^k) \to (\bar{x}, \bar{y}, 0, 0, 0, 0) \quad \text{as } k \to \infty \tag{19}$$

with

$$I_1 = I^G(x^k, y^k, \delta^k) := \{i \in \{1, \ldots, p\} \mid G_i(x^k, y^k) = \delta_j^k\},$$
$$I_2 = I^{\hat{g}}(x^k, y^k, \hat{\delta}^k) := \{j \in \{1, \ldots, r\} \mid \hat{g}_j(x^k, y^k) = \hat{\delta}_j^k\},$$
$$I_3 = I^{\hat{f} - \varphi}(x^k, y^k, \gamma_k).$$

In analogy to Remark 2, we assume without loss of generality that the sets $I^G(x^k, y^k, \delta^k)$, $I^{\hat{g}}(x^k, y^k, \hat{\delta}^k)$ and $I^{\hat{f} - \varphi}(x^k, y^k, \gamma_k)$ occurring in Definition 5.6 (ii) are constant for all $k \in \mathbb{N}$, which is reasonable as it is always possible to consider a suitable infinite subsequence. Similarly to Definition 3.5 (ii), Definition 5.6 (ii) requires that $(\bar{x}, \bar{y})$ is an AKKT point of (LLVFR) whereas part (i) can be interpreted as a weakened MFCQ at $(\bar{x}, \bar{y})$. If $(\bar{x}, \bar{y})$ is a local minimizer, there always exists at least one sequence that is a suitable candidate for Definition 5.6 (ii), as a local minimizer of (LLVFR) is always an AKKT point by Theorem 3.4. The next theorem is a direct consequence of Theorem 3.6.

Theorem 5.7. *If $(\bar{x}, \bar{y})$ is feasible for* (LLVFR) *and LLVFR-subMFC holds at* $(\bar{x}, \bar{y})$, *then* $(\bar{x}, \bar{y})$ *is a KKT point of* (LLVFR).

Finally, combining Proposition 5.1 and Theorem 5.7 directly yields the desired result.

Corollary 5.8. *Let $\{\gamma_k\}_{k=1}^\infty$ be a sequence of positive reals such that $\gamma_k \downarrow 0$ as $k \to \infty$. Assume that the point (x^k, y^k) is a KKT point of* (LLVFR$_{\gamma_k}$) *for each $k \in \mathbb{N}$. Then, any accumulation point of $\{(x^k, y^k)\}_{k=1}^\infty$ that fulfills LLVFR-subMFC is a KKT point of* (LLVFR).

5.1. Comparison to partial calmness

Partial calmness was introduced in [47] and has emerged as a standard condition to derive KKT-type optimality conditions for bilevel optimization problems which are treated by the lower-level value function reformulation [14–16, 19, 47].

Definition 5.9. Let $(\bar{x}, \bar{y})$ solve (LLVFR). Then, (LLVFR) is said to be *partially calm* at $(\bar{x}, \bar{y})$ if there exist constants $r > 0$ and $\mu > 0$ such that, for all $\alpha \in [-r, r]$ and for all $(x, y) \in B((\bar{x}, \bar{y}), r)$ with

$$G(x, y) \le 0, \quad \hat{g}(x, y) \le 0, \quad \hat{f}(x, y) - \varphi(x) + \alpha \le 0,$$

it follows that

$$F(x, y) - F(\bar{x}, \bar{y}) + \mu|\alpha| \ge 0.$$

Note that the definition of partial calmness can also be stated for $(\bar{x}, \bar{y})$ being a local minimizer of (LLVFR). Unfortunately, partial calmness is a condition that only holds under restrictive assumptions, for example when the lower-level problem is fully linear and no constraints occur in the upper level [47, Proposition 4.1]. Especially, partial calmness does not necessarily hold as soon as the lower-level problem is linear only with respect to variable y [32], see Example 5.10 below. In comparison, on the one hand, LLVFR-subMFC holds for fully linear lower-level problems (possibly with additional concave coupled upper level constraints) as well. Indeed, in this case the lower-level value function is convex [41, Lemma 2.1], and thus reformulation (LLVFR) has only concave inequality constraints. Hence, by Theorem 4.7 and the fact that concave inequality constraints imply quasinormality [45, Proposition 3], LLVFR-subMFC automatically holds. On the other hand, we have seen in Section 4 that in situations with even weaker assumptions than concavity of the inequalities, subMFC may hold as well, such that LLVFR-subMFC can be regarded as a promising candidate for an alternative to partial calmness in bilevel optimization. In particular, in the following example from [32], LLVFR-subMFC is satisfied whereas partial calmness does not hold.

Example 5.10 (Partial calmness does not imply LLVFR-subMFC)**.** Consider the following instance of an optimistic bilevel optimization problem (BP):

$$\min_{x, y=(y_1, y_2)} F(x, y) := xy_1$$
$$\text{s.t. } y \in S(x) := \operatorname*{argmin}_{y} \{\hat{f}(x, y) := -x^2 y_2 \mid \hat{g}_1(x, y) := y_2 \le 0,$$
$$\hat{g}_2(x, y) := -xy_1 + y_2 \le 0\}.$$

As shown in [32], the point $(\bar{x}, \bar{y}) := (0, (0, 0))$ is a global minimizer of the problem and the lower-level value function φ is given by $\varphi(x) = 0$ for all $x \in \mathbb{R}$. Further, it can be verified that this bilevel optimization problem is not partially calm at $(\bar{x}, \bar{y})$ [32]. In contrast, we now show that LLVFR-subMFC holds at $(\bar{x}, \bar{y})$. Concerning part (ii) of LLVFR-subMFC, the conditions (15)–(19) for this problem require $(x^k, y^k, \hat{\delta}^k, \gamma_k, \epsilon^k) \to (\bar{x}, \bar{y}, 0, 0, 0)$ as $k \to \infty$

and, for all $k \in \mathbb{N}$,

$$
\epsilon^k = \begin{pmatrix} y_1^k \\ x^k \\ 0 \end{pmatrix} + v_1^k \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} + v_2^k \begin{pmatrix} -y_1^k \\ -x^k \\ 1 \end{pmatrix} + w^k \begin{pmatrix} -2x^k y_2^k \\ 0 \\ -(x^k)^2 \end{pmatrix},
$$

$$
\begin{aligned}
v_1^k \geq 0, && y_2^k \leq \hat{\delta}_1^k, && v_1^k(y_2^k - \hat{\delta}_1^k) = 0, \\
v_2^k \geq 0, && -x^k y_1^k + y_2^k \leq \hat{\delta}_2^k, && v_2^k(-x^k y_1^k + y_2^k - \hat{\delta}_2^k) = 0, \\
w^k \geq 0, && -(x^k)^2 y_2^k \leq \gamma_k, && w^k(-(x^k)^2 y_2^k - \gamma_k) = 0.
\end{aligned}
$$

It can be checked that the sequence

$$
\{(x^k, y^k, v^k, w^k, \hat{\delta}^k, \gamma_k, \epsilon^k)\}_{k=1}^{\infty} := \left\{ \left(0, \left(0, -\frac{1}{k} \right), (0,0), 0, (0,0), \frac{1}{k}, (0,0,0) \right) \right\}_{k=1}^{\infty}
$$

is feasible for both (LLVFR) and (LLVFR$_\gamma$) and fulfills the above system. As no upper-level constraint exists, $I^{\hat{g}}(x^k, y^k, \hat{\delta}^k) = \emptyset$ and $I^{\hat{f}-\varphi}(x^k, y^k, \gamma_k) = \emptyset$, both conditions (i) and (ii) of LLVFR-subMFC are satisfied when setting $I_1 := \emptyset$, $I_2 := \emptyset$ and $I_3 := \emptyset$.

6. Final remarks

We introduced subMFC as a problem qualification for Lipschitzian optimization problems which ensures that a feasible point is a KKT point. A comparison with several CQs and the problem qualification calmness revealed that subMFC is a quite weak condition, as it does not imply any of these conditions while being implied by quasinormality. Further, we demonstrated the application of subMFC to bilevel optimization problems and derived the tailored version LLVFR-subMFC that does not imply the well-known partial calmness condition. In our future research we aim at applying our findings to different problem classes like MPCCs. Moreover, from a practical standpoint, it can be promising to focus on exploiting our approach when designing numerical algorithms for the treatment of problems belonging to the aforementioned problem classes.

Disclosure of interest

The authors report there are no competing interests to declare.

References

[1] Achtziger, W., Kanzow, C.: Mathematical programs with vanishing constraints: Optimality conditions and constraint qualifications. Math. Program. **114**(1), 69–99 (2008).

[2] Andreani, R., Haeser, G., Martínez, J.M.: On sequential optimality conditions for smooth constrained optimization. Optim. **60**(5), 627–641 (2011).

[3] Andreani, R., Haeser, G., Schuverdt, M.L., Silva, P.J.S.: A relaxed constant positive linear dependence constraint constraint qualification and applications. Math. Program. **135**(1), 255–273 (2012).

[4] Andreani, R., Haeser, G., Schuverdt, M.L., Silva, P.J.S.: Two new weak constraint qualifications and applications. SIAM J. Optim. **22**(3), 1109–1135 (2012).

[5] Andreani, R., Haeser, G., Secchin, L.D., Silva, P.J.S.: New sequential optimality conditions for mathematical programs with complementarity constraints and algorithmic consequences. SIAM J. Optim. **29**(4), 3201–3230 (2019).

[6] Andreani, R., Martínez, J.M., Ramos, A., Silva, P.J.S.: A cone-continuity constraint qualification and algorithmic consequences. SIAM J. Optim. **26**(1), 96–110 (2016).

[7] Andreani, R., Martínez, J.M., Schuverdt, M.L.: On the relation between constant positive linear dependence condition and quasinormality constraint qualification. J. Optim. Theory Appl. **125**(2), 473–485 (2005).

[8] Andreani, R., Martínez, J.M., Svaiter, B.F.: A new sequential optimality condition for constrained optimization and algorithmic consequences. SIAM J. Optim. **20**(6), 3533–3554 (2010).

[9] Bertsekas, D.P.: Nonlinear Programming. Optimization and Computation Series. Athena Scientific (1999)

[10] Bertsekas, D.P., Ozdaglar, A.E.: Pseudonormality and a Lagrange multiplier theory for constrained optimization. J. Optim. Theory Appl. **114**(2), 287–343 (2002).

[11] Birgin, E.G., Martínez, J.M.: Practical Augmented Lagrangian Methods for Constrained Optimization. Fundamentals of Algorithms. Society for Industrial and Applied Mathematics (2014)

[12] Bueno, L.F., Friedlander, A., Martínez, J.M., Sobral, F.N.C.: Inexact restoration method for derivative-free optimization with smooth constraints. SIAM J. Optim. **23**(2), 1189–1213 (2013).

[13] Clarke, F.H.: Optimization and Nonsmooth Analysis. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics (1990)

[14] Dempe, S., Dutta, J., Mordukhovich, B.S.: New necessary optimality conditions in optimistic bilevel programming. Optim. **56**(5-6), 577–604 (2007).

[15] Dempe, S., Zemkoho, A.B.: The generalized Mangasarian-Fromowitz constraint qualification and optimality conditions for bilevel programs. J. Optim. Theory Appl. **148**(1), 46–68 (2011).

[16] Dempe, S., Zemkoho, A.B.: The bilevel programming problem: reformulations, constraint qualifications and optimality conditions. Math. Program. **138**(1), 447–473 (2013).

[17] Fischer, A., Friedlander, A.: A new line search inexact restoration approach for nonlinear programming. Comput. Optim. Appl. **46**, 333–346 (2010).

[18] Fischer, A., Izmailov, A.F., Jelitte, M.: Constrained Lipschitzian error bounds and noncritical solutions of constrained equations. Set-Valued Var. Anal. **29**(3), 745–765 (2021).

[19] Fischer, A., Zemkoho, A.B., Zhou, S.: Semismooth Newton-type method for bilevel optimization: global convergence and extensive numerical experiments. Optim. Methods Softw. **37**(5), 1770–1804 (2022).

[20] Guo, L., Ye, J.J., Zhang, J.: Mathematical programs with geometric constraints in Banach spaces: Enhanced optimality, exact penalty, and sensitivity. SIAM J. Optim. **23**(4), 2295–2319 (2013).

[21] Hestenes, M.R.: Optimization Theory: The Finite Dimensional Case. Pure & Applied Mathematics S. Wiley (1975)

[22] Janin, R.: Directional derivative of the marginal function in nonlinear programming. In: A.V. Fiacco (ed.) Sensitivity, Stability and Parametric Analysis, pp. 110–126. Springer (1984).

[23] Kyparisis, J.: On uniqueness of Kuhn-Tucker multipliers in nonlinear programming. Math. Program. **32**(2), 242–246 (1985).

[24] Lampariello, L., Sagratella, S.: Numerically tractable optimistic bilevel problems. Comput.

Optim. Appl. **76**(2), 277–303 (2020).

[25] Li, W.: Abadie's constraint qualification, metric regularity, and error bounds for differentiable convex inequalities. SIAM J. Optim. **7**(4), 966–978 (1997).

[26] Lin, G.H., Xu, M., Ye, J.J.: On solving simple bilevel programs with a nonconvex lower level program. Math. Program. **144**(1), 277–305 (2014).

[27] Luo, Z.Q., Pang, J.S., Ralph, D.: Mathematical Programs with Equilibrium Constraints. Cambridge University Press (1996).

[28] Martínez, J.M., Svaiter, B.F.: A practical optimality condition without constraint qualifications for nonlinear programming. J. Optim. Theory Appl. **118**(1), 117–133 (2003).

[29] Mehlitz, P.: Asymptotic stationarity and regularity for nonsmooth optimization problems. J. Nonsmooth Anal. Optim. **1**, 6575 (2020).

[30] Mehlitz, P.: Stationarity conditions and constraint qualifications for mathematical programs with switching constraints. Math. Program. **181**(1), 149–186 (2020).

[31] Mehlitz, P.: Asymptotic regularity for Lipschitzian nonlinear optimization problems with applications to complementarity constrained and bilevel programming. Optim. **72**(1), 277–320 (2023).

[32] Mehlitz, P., Minchenko, L., Zemkoho, A.B.: A note on partial calmness for bilevel optimization problems with linearly structured lower level. Optim. Lett. **15**(4), 1277–1291 (2021).

[33] Minchenko, L., Stakhovski, S.: On relaxed constant rank regularity condition in mathematical programming. Optim. **60**(4), 429–440 (2011).

[34] Minchenko, L., Tarakanov, A.: On error bounds for quasinormal programs. J. Optim. Theory Appl. **148**(3), 571–579 (2011).

[35] Mordukhovich, B.S.: Variational Analysis and Applications. Springer Monographs in Mathematics. Springer (2018).

[36] Qi, L., Wei, Z.: On the constant positive linear dependence condition and its application to SQP methods. SIAM J. Optim. **10**(4), 963–981 (2000).

[37] Ramos, A.: Mathematical programs with equilibrium constraints: a sequential optimality condition, new constraint qualifications and algorithmic consequences. Optim. Methods Softw. **36**(1), 45–81 (2021).

[38] Rimpi, Lalitha, C.S.: Constraint qualifications in nonsmooth optimization: Classification and inter-relations. J. Nonlinear Var. Anal. **6**, 83–99 (2022).

[39] Rimpi, Lalitha, C.S.: Constraint qualifications in terms of convexificators for nonsmooth programming problems with mixed constraints. Optim. **72**(8), 2019–2038 (2023).

[40] Rockafellar, R.T., Wets, R.J.B.: Variational Analysis. Grundlagen der mathematischen Wissenschaften. Springer (1998)

[41] Tanino, T., Ogawa, T.: An algorithm for solving two-level convex optimization problems. Int. J. Syst. Sci. **15**(2), 163–174 (1984).

[42] Wachsmuth, G.: On LICQ and the uniqueness of Lagrange multipliers. Operations Research Letters **41**(1), 78–80 (2013).

[43] Xu, M., Ye, J.J.: Relaxed constant positive linear dependence constraint qualification and its application to bilevel programs. J. Global Optim. **78**(1), 181–205 (2020).

[44] Ye, J.J., Yuan, X., Zeng, S., Zhang, J.: Difference of convex algorithms for bilevel programs with applications in hyperparameter selection. Math. Program. **198**(2), 1583–1616 (2023).

[45] Ye, J.J., Zhang, J.: Enhanced Karush–Kuhn–Tucker condition and weaker constraint qualifications. Math. Program. **139**(1), 353–381 (2013).

[46] Ye, J.J., Zhu, D.: New necessary optimality conditions for bilevel programs by combining the MPEC and value function approaches. SIAM J. Optim. **20**(4), 1885–1905 (2010).

[47] Ye, J.J., Zhu, D.L.: Optimality conditions for bilevel programming problems. Optim. **33**(1), 9–27 (1995).

Appendix A. Examples relating subMFC to other conditions

A.1. SubMFC and the local error bound condition are independent conditions

By the following example, it is first shown that subMFC can hold even if the local error bound condition is violated.

Example A.1. Recall the optimization problem from Example 4.8. The feasible set $D = \{(x_1, x_2) \in \mathbb{R}^2 \mid g_1(x_1, x_2) \leq 0\}$ of this problem is visualized in Figure A1.
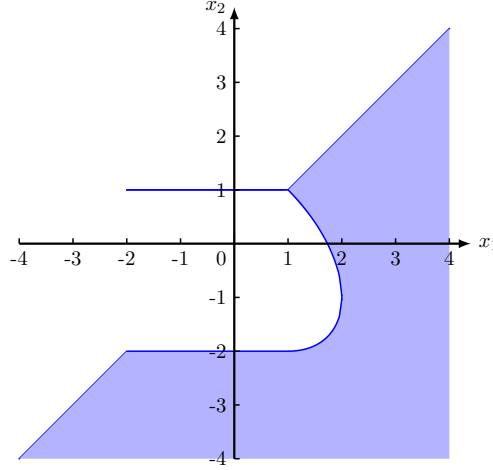

Figure A1. The feasible set D of the optimization problem in Example A.1 is depicted in blue.

We have seen that subMFC holds at the global minimizer $\bar{x} = (\bar{x}_1, \bar{x}_2) = (-2, 1)$. However, the local error bound condition does not hold at $\bar{x}$, i.e., we cannot find $c > 0$ and $r > 0$ such that

$$d_D(x) \leq cg_{1+}(x) \quad \forall x = (x_1, x_2) \in B((\bar{x}_1, \bar{x}_2), r).$$

To prove this claim, assume that some $r > 0$ is fixed. Then, for $x = (x_1, x_2) = (-2, 1 + \tau)$ with $\tau \in [0, r]$ we have $x \in B((\bar{x}_1, \bar{x}_2), r)$. We obtain $d_D(x) = \tau$, as the point in D that is closest to x is $(-2, 1)$, see Figure A1. Further, we have

$$g_{1+}(x) = \max\{0, (1 + \tau)^3 - 3(1 + \tau) + 2\} = \tau^3 + 3\tau^2.$$

Thus, for the local error bound condition to hold, a constant $c > 0$ would be needed such that $\tau \leq c(\tau^3 + 3\tau^2)$ is satisfied for all $\tau \in [0, r]$. Obviously, such a constant c does not exist. As $r > 0$ was chosen arbitrarily, the local error bound condition cannot hold at $\bar{x}$.

On the other hand, we can also show that the local error bound condition can hold even if subMFC is violated, meaning that both conditions are independent of each other.

Example A.2. Let us again consider the optimization problem from Example 3.7. There, we have seen that subMFC is not valid at the local minimizer $\bar{x} = (\bar{x}_1, \bar{x}_2) = (0, 0)$. In order to show that the local error bound condition holds at $\bar{x}$, we rely on the sufficient condition from [18, Corollary 2]. To apply this condition, we define the function $F(u) := g_+(x)$ for which we have $F^{-1}(0) = \bar{x}$ and $\widehat{T}_{F^{-1}(0)}(\bar{x}) = \{(0, 0)\} = \{d \in \mathbb{R}^2 \mid F'(\bar{x}; d) = 0\}$, where $\widehat{T}_{F^{-1}(0)}(\bar{x})$ is a subset of the contingent cone defined in Section 4.6; see [18] for details. Thus, all requirements of [18, Corollary 2] are met and the local error bound condition is satisfied at $\bar{x}$.

30

A.2. SubMFC and the cone-continuity property are independent conditions

The next example demonstrates that CCP can be violated while subMFC holds.

Example A.3. Consider the following instance of optimization problem (P):

$$\min\ (x_1 + x_2)x_1 \quad \text{s.t.} \quad \begin{aligned} g_1(x_1, x_2) &:= x_1 x_2 \le 0, \\ g_2(x_1, x_2) &:= -x_1 x_2 \le 0, \\ g_3(x_1, x_2) &:= -x_1 \le 0. \end{aligned}$$

By the first two constraints, x_1 or x_2 must be equal to 0 for any feasible point. In the case $x_1 = 0$, the objective function value becomes $f(0, x_2) = 0$, while in the case $x_2 = 0$, the objective function value is nonnegative, as $f(x_1, 0) = x_1^2$. Therefore, all points $(0, x_2)$ with $x_2 \in \mathbb{R}$ are global solutions. We will particularly look at the point $\bar{x} = (\bar{x}_1, \bar{x}_2) = (0, 1)$.

(1) First, we show that CCP is violated at $\bar{x}$. For this purpose, due to $I(\bar{x}) = \{1, 2, 3\}$, the set $K(\bar{x})$ can be computed as

$$\begin{aligned} K(\bar{x}) &= \left\{ \sum_{j \in I(\bar{x})} v_j \nabla g_j(\bar{x}) \mid v_j \ge 0 \right\} \\ &= \left\{ v_1 \begin{pmatrix} \bar{x}_2 \\ \bar{x}_1 \end{pmatrix} + v_2 \begin{pmatrix} -\bar{x}_2 \\ -\bar{x}_1 \end{pmatrix} + v_3 \begin{pmatrix} -1 \\ 0 \end{pmatrix} \mid v_1, v_2, v_3 \ge 0 \right\} \\ &= \left\{ \begin{pmatrix} v_1 - v_2 - v_3 \\ 0 \end{pmatrix} \mid v_1, v_2, v_3 \ge 0 \right\} \\ &= \mathbb{R} \times \{0\}. \end{aligned}$$

For $x^k = (x_1^k, x_2^k) = \left(\frac{1}{k}, 1\right)$ with $x^k \to \bar{x}$ and $(v_1, v_2, v_3) = (2k, k, k)$ for $k > 0$, we get

$$w^k := \sum_{j \in I(\bar{x})} v_j \nabla g_j(x^k) = 2k \begin{pmatrix} 1 \\ \frac{1}{k} \end{pmatrix} + k \begin{pmatrix} -1 \\ -\frac{1}{k} \end{pmatrix} + k \begin{pmatrix} -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Thus, we obtain $w^k \in K(x^k)$ and $w^k \to \bar{w} := (0, 1) \in \limsup_{x^k \to \bar{x}} K(x^k)$. However, $\bar{w} \notin K(\bar{x})$, such that inclusion (13) does not hold. Hence, CCP is violated at $\bar{x}$.

(2) In contrast, subMFC is fulfilled at $\bar{x}$. To see this, consider the AKKT conditions in $\bar{x}$, which require $(x^k, \delta^k, \epsilon^k) \to (\bar{x}, 0, 0)$ as $k \to \infty$ and

$$\begin{pmatrix} \epsilon_1^k \\ \epsilon_2^k \end{pmatrix} = \begin{pmatrix} 2x_1^k + x_2^k \\ x_1^k \end{pmatrix} + u_1^k \begin{pmatrix} x_2^k \\ x_1^k \end{pmatrix} + u_2^k \begin{pmatrix} -x_2^k \\ -x_1^k \end{pmatrix} + u_3^k \begin{pmatrix} -1 \\ 0 \end{pmatrix},$$

$$\begin{aligned} u_1^k &\ge 0, & x_1^k x_2^k &\le \delta_1^k, & u_1^k(x_1^k x_2^k - \delta_1^k) &= 0, \\ u_2^k &\ge 0, & -x_1^k x_2^k &\le \delta_2^k, & u_2^k(-x_1^k x_2^k - \delta_2^k) &= 0, \\ u_3^k &\ge 0, & -x_1^k &\le \delta_3^k, & u_3^k(-x_1^k - \delta_3^k) &= 0 \end{aligned}$$

for all $k \in \mathbb{N}$. One sequence to fulfill the above conditions is given by

$$\{(x^k, u^k, \delta^k, \epsilon^k)\}_{k=1}^{\infty} := \left\{ \left(\left(\frac{1}{k}, 1\right), \left(0, 1, \frac{2}{k}\right), \left(\frac{2}{k}, -\frac{1}{k}, -\frac{1}{k}\right), (0, 0) \right) \right\}_{k=1}^{\infty}.$$

For this sequence it holds that $I(x^k, \delta^k) = \{2, 3\}$ for all $k \in \mathbb{N}$. Hence, with $I := \{2, 3\}$,

$$0 \neq \sum_{j \in I} v_j \nabla g_j(\bar{x}) = v_2 \begin{pmatrix} -\bar{x}_2 \\ -\bar{x}_1 \end{pmatrix} + v_3 \begin{pmatrix} -1 \\ 0 \end{pmatrix} = v_2 \begin{pmatrix} -1 \\ 0 \end{pmatrix} + v_3 \begin{pmatrix} -1 \\ 0 \end{pmatrix}$$

holds for all $(v_2, v_3) \neq (0,0)$ with $v_2 \geq 0, v_3 \geq 0$, i.e., subMFC is satisfied at $\bar{x}$.

Returning to a known example, we can also show the opposite, i.e., that CCP can be valid even if subMFC is violated.

Example A.4. We again consider Example 3.7, where subMFC is violated at the local minimizer $\bar{x} = (\bar{x}_1, \bar{x}_2) = (0,0)$. To verify CCP, we compute $I(\bar{x}) = \{1, 2, 3, 4\}$ and

$$K(\bar{x}) = \left\{ \sum_{j \in I(\bar{x})} v_j \partial g_j(\bar{x}) \mid v_j \geq 0 \right\}$$
$$= \left\{ v_1 \begin{pmatrix} [-1, 1] \\ -1 \end{pmatrix} + v_2 \begin{pmatrix} [-1, 1] \\ 1 \end{pmatrix} + v_3 \begin{pmatrix} 0 \\ 2 \end{pmatrix} + v_4 \begin{pmatrix} 0 \\ -2 \end{pmatrix} \mid v_1, v_2, v_3, v_4 \geq 0 \right\} = \mathbb{R}^2.$$

Thus, CCP obviously holds at $\bar{x}$.

A.3. SubMFC and the Guignard constraint qualification are independent conditions

We now show that subMFC can hold even if both ACQ and GCQ are violated.

Example A.5. Consider the following instance of optimization problem (P):

$$\min x_2 \quad \text{s.t.} \quad \begin{aligned} g_1(x_1, x_2) &\coloneqq x_1 x_2 \leq 0, \\ g_2(x_1, x_2) &\coloneqq -x_1 x_2 \leq 0, \\ g_3(x_1, x_2) &\coloneqq x_1^2 - x_2 \leq 0. \end{aligned}$$

The feasible set is given by $D = \{0\} \times [0, \infty)$. Hence, the global solution to the problem is $\bar{x} = (\bar{x}_1, \bar{x}_2) = (0,0)$.

(1) First, we show that both ACQ and GCQ are violated at $\bar{x}$. The computation of the contingent cone and the cone of locally constrained directions to D at $\bar{x}$ yields

$$T_D(\bar{x}) = \left\{ \begin{pmatrix} 0 \\ d_2 \end{pmatrix} \in \mathbb{R}^2 \mid d_2 \geq 0 \right\} = D, \quad L_D(\bar{x}) = \left\{ \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} \in \mathbb{R}^2 \mid d_1 \in \mathbb{R}, d_2 \geq 0 \right\},$$

such that ACQ is violated at $\bar{x}$ due to $L_D(\bar{x}) \nsubseteq T_D(\bar{x})$. Due to $\operatorname{cl} \operatorname{conv} T_D(\bar{x}) = T_D(\bar{x})$, GCQ does not hold at $\bar{x}$ as well.

(2) Nevertheless, subMFC holds at $\bar{x}$. To this end, consider the AKKT conditions in $\bar{x}$, which consist of $(x^k, \delta^k, \epsilon^k) \to (\bar{x}, 0, 0)$ as $k \to \infty$ and

$$\begin{pmatrix} \epsilon_1^k \\ \epsilon_2^k \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} + u_1^k \begin{pmatrix} x_2^k \\ x_1^k \end{pmatrix} + u_2^k \begin{pmatrix} -x_2^k \\ -x_1^k \end{pmatrix} + u_3^k \begin{pmatrix} 2x_1^k \\ -1 \end{pmatrix},$$
$$u_1^k \geq 0, \qquad x_1^k x_2^k \leq \delta_1^k, \qquad u_1^k(x_1^k x_2^k - \delta_1^k) = 0,$$
$$u_2^k \geq 0, \qquad -x_1^k x_2^k \leq \delta_2^k, \qquad u_2^k(-x_1^k x_2^k - \delta_2^k) = 0,$$
$$u_3^k \geq 0, \quad (x_1^k)^2 - x_2^k \leq \delta_3^k, \quad u_3^k((x_1^k)^2 - x_2^k - \delta_3^k) = 0$$

for all $k \in \mathbb{N}$. Clearly, the sequence

$$\{(x^k, u^k, \delta^k, \epsilon^k)\}_{k=1}^\infty := \left\{ (0,0), (0,0,1), \left(\frac{1}{k}, \frac{1}{k}, 0\right), (0,0) \right\}_{k=1}^\infty$$

fulfills the above AKKT conditions. For this sequence we have $I(x^k, \delta^k) = \{3\}$ for all $k \in \mathbb{N}$. Thus, setting $I := \{3\}$, it follows that

$$0 \neq \sum_{j \in I} v_j \nabla g_j(\bar{x}) = v_3 \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$

is fulfilled for all $v_3 > 0$. Hence, subMFC is satisfied at $\bar{x}$.

Again revisiting Example 3.7, we see that ACQ and GCQ can be satisfied even if subMFC is not.

Example A.6. Recall again the optimization problem from Example 3.7 where we have seen that subMFC does not hold at the local minimizer $\bar{x} = (\bar{x}_1, \bar{x}_2) = (0,0)$. The computation of the contingent cone and the outer linearization cone to D at $\bar{x}$ yields

$$T_D(\bar{x}) = \{0\} = L_D(\bar{x}),$$

such that, due to $\operatorname{cl} \operatorname{conv} T_D(\bar{x}) = T_D(\bar{x})$, both ACQ and GCQ are fulfilled.

A.4. SubMFC does not imply calmness in the sense of Clarke

Finally, the following example demonstrates that subMFC can hold even if calmness in the sense of Clarke does not.

Example A.7. Consider the following instance of optimization problem (P):

$$\min f(x_1, x_2) := \ln(x_1) - 2x_1 + x_2$$
$$\text{s.t. } g_1(x_1, x_2) := 2x_1 - x_2^3 + 9x_2^2 - 29x_2 + 31 \leq 0,$$
$$g_2(x_1, x_2) := -\frac{1}{2}x_1 + \frac{1}{54}x_2^3 - \frac{1}{6}x_2^2 + x_2 - \frac{3}{2} \leq 0,$$
$$g_3(x_1, x_2) := x_2 - 3 \leq 0.$$

(1) We first show that $\bar{x} = (\bar{x}_1, \bar{x}_2) = (1,3)$ is a global solution to the problem. For this purpose, note that g_1 and g_2 provide an upper and a lower bound for x_1. Indeed, for any feasible point $x = (x_1, x_2)$, it is required that

$$\frac{1}{27}x_2^3 - \frac{1}{3}x_2^2 + 2x_2 - 3 \leq (x_1 \leq) \frac{1}{2}x_2^3 - \frac{9}{2}x_2^2 + \frac{29}{2}x_2 - \frac{31}{2}$$

holds, which is equivalent to

$$\frac{25}{54}\left(x_2^3 - 9x_2^2 + 27x_2 - 27\right) \geq 0 \iff \frac{25}{54}\left(x_2 - 3\right)^3 \geq 0 \iff x_2 - 3 \geq 0.$$

Together with $g_3(x_1, x_2) \leq 0$, this yields $x_2 = 3$ for any feasible point. In turn, g_1 and g_2 for $x_2 = 3$ read as

$$g_1(x_1, 3) = 2x_1 - 2 \quad \text{and} \quad g_2(x_1, 3) = -\frac{1}{2}x_1 + \frac{1}{2}.$$

Thus, it can be verified that the point $\bar{x} = (1,3)$ is the only feasible point and therefore a global solution to the problem with $f(1,3) = 1$.

(2) We now demonstrate that calmness is violated at $\bar{x} = (1,3)$. To this end, let us suppose the contrary, i.e., that calmness holds at the solution $\bar{x} = (1,3)$. Then, there exist $r > 0$ and $\mu > 0$ such that for all $\alpha \in B(0, r)$ it holds that

$$\ln(x_1) - 2x_1 + x_2 - 1 + \mu \|\alpha\| \geq 0 \quad \forall x \in \{(x_1, x_2) \in B(\bar{x}, r) \mid g(x_1, x_2) + \alpha \leq 0\}. \quad \text{(A1)}$$

Consider the sequence $\{(\alpha^k, x^k)\}_{k=1}^{\infty}$ with

$$(\alpha^k, x^k) := \left(\left(-\frac{1}{k^3}, 0, 0 \right), \left(1 - \frac{1}{k}, 3 - \frac{1}{k} \right) \right)$$

which fulfills $(\alpha^k, x^k) \to ((0,0,0), (\bar{x}_1, \bar{x}_2))$ as $k \to \infty$ and

$$g_1(x_1^k, x_2^k) + \alpha_1^k = 0, \quad g_2(x_1^k, x_2^k) + \alpha_2^k = -\frac{1}{54k^3} < 0, \quad g_3(x_1^k, x_2^k) + \alpha_3^k = -\frac{1}{k} < 0.$$

However, as $\|\alpha^k\| = \frac{1}{k^3}$, it follows from (A1) that

$$\ln(x_1^k) - 2x_1^k + x_2^k - 1 + \mu \|\alpha^k\| = \frac{1}{k} + \ln \left(1 - \frac{1}{k} \right) + \frac{\mu}{k^3} \geq 0.$$

This inequality is equivalent to

$$\mu \geq -k^2 \left(k \ln \left(1 - \frac{1}{k} \right) + 1 \right), \quad \text{(A2)}$$

and it can be shown that the right-hand side of (A2) converges to ∞ for $k \to \infty$. Thus, we cannot find a finite $\mu > 0$ to fulfill the inequality, and hence, the problem is not calm at $\bar{x}$.

(3) In contrast, subMFC is fulfilled at $\bar{x}$. To verify this claim, we take a look at the AKKT conditions in $\bar{x}$, i.e., we need to satisfy $(x^k, \delta^k, \epsilon^k) \to (\bar{x}, 0, 0)$ as $k \to \infty$ and

$$\begin{pmatrix} \epsilon_1^k \\ \epsilon_2^k \end{pmatrix} = \begin{pmatrix} \frac{1}{x_1^k} - 2 \\ 1 \end{pmatrix} + u_1^k \begin{pmatrix} 2 \\ -3(x_2^k)^2 + 18x_2^k - 29 \end{pmatrix} + u_2^k \begin{pmatrix} -\frac{1}{2} \\ \frac{1}{18}(x_2^k)^2 - \frac{1}{3}x_2^k + 1 \end{pmatrix} + u_3^k \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

$$u_i^k \geq 0, \quad g_i(x_1^k, x_2^k) \leq \delta_i^k, \quad u_i^k(g_i(x_1^k, x_2^k) - \delta_i^k) = 0, \quad \forall i = 1, \ldots, 3$$

for all $k \in \mathbb{N}$. Consider the sequence

$$\left\{ (x^k, u^k, \delta^k, \epsilon^k) \right\}_{k=1}^{\infty} := \left\{ \left(\left(1 - \frac{1}{k}, 3 - \frac{1}{k} \right), \left(\frac{k-2}{2k-2}, 0, 0 \right), \left(\frac{1}{k^3}, 0, 0 \right), \left(0, \epsilon_2^k \right) \right) \right\}_{k=1}^{\infty}$$

with $\epsilon_2^k := \frac{2k^2 - 3k + 6}{2(k-1)k^2} \to 0$ as $k \to \infty$. Then, the first line of the AKKT system reads as

$$\begin{pmatrix} 0 \\ \epsilon_2^k \end{pmatrix} = \begin{pmatrix} \frac{k}{k-1} - 2 \\ 1 \end{pmatrix} + \frac{k-2}{2k-2} \cdot \begin{pmatrix} 2 \\ -\frac{3}{k^2} - 2 \end{pmatrix},$$

while the other AKKT conditions are fulfilled as well due to $(x^k, \delta^k, \epsilon^k) \to (\bar{x}, 0, 0)$ as $k \to \infty$ and

$$g_1(x_1^k, x_2^k) = \frac{1}{k^3} = \delta_1^k, \quad g_2(x_1^k, x_2^k) = -\frac{1}{54k^3} < 0 = \delta_2^k, \quad g_3(x_1^k, x_2^k) = -\frac{1}{k} < 0 = \delta_3^k.$$

Thus, for the sequence $\{(x^k, u^k, \delta^k, \epsilon^k)\}_{k=1}^{\infty}$, it holds for all $k \in \mathbb{N}$ that $I(x^k, \delta^k) = \{1\}$. Further, with $I := \{1\}$, condition (i) of subMFC means that

$$0 \neq \sum_{j \in I} v_j \nabla g_j(\bar{x}) = v_1 \begin{pmatrix} 2 \\ -3\bar{x}_2^2 + 18\bar{x}_2 - 29 \end{pmatrix} = v_1 \begin{pmatrix} 2 \\ -2 \end{pmatrix}$$

has to be fulfilled for all $v_1 > 0$, and this is indeed the case. Thus, subMFC holds at $\bar{x}$.

Appendix B. Examples and results verifying relations between different conditions

In this section, we briefly summarize how the relations depicted in Figure 1, which have not yet been discussed in the course of this paper, can be verified. Results concerning implications can be obtained as follows:

- The implication between LICQ and MFCQ follows trivially from the respective definitions, as LICQ at a feasible point $\bar{x}$ of (P) demands the linear independence of the subgradients of all constraints that are active in $\bar{x}$, whereas for MFCQ only their positive linear independence is required. Analogously, the implications between LICQ and SMFC as well as SMFC and MFCQ are obtained similarly to the smooth case.
- The nonsmooth extension of CRCQ is given in [43] and trivially yields that CRCQ is implied by LICQ.
- In [22], it is shown that CRCQ and MFCQ do not imply each other.

Remaining results concerning the strictness of the presented implications can be established using the following examples:

- The feasible set given by $g_1(x) := 2x \leq 0$ and $g_2(x) := x \leq 0$ and the feasible point $\bar{x} = 0$ can be used to see that the implications between LICQ and MFCQ as well as LICQ and CRCQ are strict.
- The optimization problem (P) with $f(x) := x^2$, $g_1(x) := x$ and $g_2(x) := -x$ with the global optimizer $\bar{x} = 0$ can be used to see that the implications between MFCQ and pseudonormality as well as MFCQ and calmness are strict.
- Further, [7, Counterexample 4.4] shows that pseudonormality can hold even if not all constraint functions are concave and that the implication between CPLD and quasinormality is strict.
- For the feasible set given by $g_1(x_1, x_2) := -x_2 e^{x_1} \leq 0$ and $g_2(x_1, x_2) := x_2 \leq 0$, which is inspired by [7], the local error bound condition is satisfied at $\bar{x} = (0, 0)$ while CPLD is not.
- When considering the feasible set given by $g_1(x_1, x_2) := |x_1| - x_2 \leq 0$ and $g_2(x_1, x_2) := x_2 \leq 0$ and the point $\bar{x} = (0, 0)$, it can be shown that the implications between quasinormality and the local error bound condition as well as between quasinormality and ACQ are strict.
- In [6], it is shown that the implication between CCP and ACQ is strict.
- The strictness of the implication between ACQ and GCQ can be verified by considering the feasible set given by $g_1(x_1, x_2) := -x_1 \leq 0$, $g_2(x_1, x_2) := -x_2 \leq 0$ and $g_3(x_1, x_2) := x_1 x_2 \leq 0$ and the feasible point $\bar{x} = (0, 0)$.