Asymptotic regularity for Lipschitzian nonlinear optimization problems with applications to complementarity constrained and bilevel programming

Patrick Mehlitz

Institute of Mathematics, Brandenburgische Technische Universität Cottbus–Senftenberg, Cottbus, Germany

ABSTRACT
Asymptotic stationarity and regularity conditions turned out to be quite useful to study the qualitative properties of numerical solution methods for standard nonlinear and complementarity constrained programs. In this paper, we first extend these notions to nonlinear optimization problems with non-smooth but Lipschitzian data functions in order to find reasonable notions of asymptotic stationarity and regularity in terms of Clarke’s and Mordukhovich’s subdifferential construction. Particularly, we compare the associated novel asymptotic constraint qualifications with already existing ones. The second part of the paper presents two applications of the obtained theory. On the one hand, we specify our findings for complementarity constrained optimization problems and recover recent results from the literature which demonstrates the power of the approach. Furthermore, we hint at potential extensions to or- and vanishing constrained optimization. On the other hand, we demonstrate the usefulness of asymptotic regularity in the context of bilevel optimization. More precisely, we justify a well-known stationarity system for affinely constrained bilevel optimization problems in a novel way. Afterwards, we suggest a solution algorithm for this class of bilevel optimization problems which combines a penalty method with ideas from DC-programming. After a brief convergence analysis, we present the results of some numerical experiments.

1. Introduction
During the last decade, asymptotic (sometimes referred to as sequential) notions of stationarity and regularity have been developed for standard nonlinear optimization problems (see, e.g. [1–5]), complementarity constrained programs (see [6,7]), cardinality constrained programs (see [8,9]), and programs in abstract

CONTACT Patrick Mehlitz mehlitz@b-tu.de
Dedicated to Stephan Dempe on the occasion of his 65th birthday.
This article has been republished with minor changes. These changes do not impact the academic content of the article.

© 2022 Informa UK Limited, trading as Taylor & Francis Group
spaces (see [10,11]). Extensions to nonsmooth optimization problems have been discussed recently in [12], based on Goldstein’s $\varepsilon$-subdifferential, and in [13], where the tools of limiting variational analysis have been exploited. The huge interest in these concepts is based on their significant relevance for the investigation of convergence properties associated with solution algorithms tailored for the aforementioned problem classes. More precisely, some numerical methods naturally produce asymptotically stationary points so the question arises which type of condition is necessary to hold at the limit in order to guarantee its stationarity in the classical sense. The resulting asymptotic regularity conditions have been shown to serve as comparatively weak constraint qualifications for a bunch of problem classes in mathematical programming.

In this paper, we apply the concepts of asymptotic stationarity and regularity to nonlinear optimization problems of the form

$$\min \{ \varphi_0(z) \mid \varphi_i(z) \leq 0 \ (i \in I), \ \varphi_i(z) = 0 \ (i \in J) \},$$

(P)

where the functions $\varphi_0, \ldots, \varphi_{p+q} : \mathbb{R}^n \to \mathbb{R}$ are assumed to be locally Lipschitz continuous but not necessarily smooth in a neighbourhood of a given reference point. Here, we use $I := \{1, \ldots, p\}$ and $J := \{p + 1, \ldots, p + q\}$. For that purpose, we will exploit the subdifferential concepts of Clarke and Mordukhovich (see e.g. [14,15]), respectively, since these generalized derivatives are outer semicontinuous (in the sense of set-valued mappings) by construction which will be beneficial for our theoretical investigations. In [12, Section 1], the authors point out that this approach may have the disadvantage that numerical methods associated with (P) may not compute stationary points in this new sense. However, as we will see in Section 4, our results recover recently introduced notions of asymptotic stationarity and regularity for so-called mathematical programs with complementarity constraints (MPCCs for short) (see [16,17]), which have been shown to be useful in numerical practice (see [6,7]). We note that our model program (P) covers other prominent classes from disjunctive programming like so-called or- and vanishing constrained programs (see [18–22] and Remark 4.3), so that our findings are likely to possess reasonable extensions to these models as well. Based on our new notions of asymptotic stationarity, we introduce three novel sequential constraint qualifications which guarantee that asymptotically stationary points of (P) are stationary in Clarke’s or Mordukhovich’s sense, i.e. that these points satisfy Karush–Kuhn–Tucker-type stationarity conditions based on Clarke’s or Mordukhovich’s subdifferential. Afterwards, we study the relationship between these new regularity conditions and already available constraint qualifications from nonsmooth programming. Particularly, we address connections to a nonsmooth version of the so-called relaxed constant positive linear dependence constraint qualification (RCPLD), which has been introduced for smooth standard nonlinear optimization problems in [23] and extended to nonsmooth programs in [24] quite recently.
As already mentioned, we apply our quite general findings regarding the abstract model (P) to MPCCs in Section 4 in order to underline the particular value as well as the applicability of these results. In Section 5, we demonstrate the power of asymptotic stationarity and regularity in the context of bilevel optimization with affine data in the upper level constraints as well as in the overall lower level. It is well known that bilevel optimization problems are notoriously difficult due to their inherent irregularity, nonsmoothness, and nonconvexity, while being a major topic in the focus of many researchers because of their overwhelming practical relevance with respect to (w.r.t.) the modelling of real-world applications from finance, economics, or natural and engineering sciences (see [25–27] for an introduction to bilevel optimization and a comprehensive literature review). Here, we make use of the so-called optimal value transformation of bilevel optimization problems in order to transfer the program of interest into the form (P). Noting that the latter is inherently asymptotically regular in the investigated setting, we are in a position to state necessary optimality conditions without any further assumptions or the use of partial penalization arguments. Afterwards, we suggest a solution method for the considered problem class which penalizes the constraint comprising the optimal value function and uses methods from DC-programming, where DC abbreviates difference of convex functions (see [28,29] for an overview), in order to solve the subproblems. As we will see, this approach is computationally reasonable since it exploits only pointwise evaluations of function values and subgradients of the optimal value function which can be easily computed while the overall optimal value function may remain an implicitly given object. Furthermore, we can apply our abstract theory from Section 3 in order to demonstrate that our method computes stationary points of the bilevel optimization problem of interest. Some numerical results visualize the computational performance of the method.

The remaining parts of this paper are organized as follows. In Section 2, we summarize the notation used in this manuscript, recall some fundamental notions from nonsmooth differentiation, and present some preliminary results. Section 3 is dedicated to the formal introduction of asymptotic stationarity and regularity notions which address (P). Noting that we proceed in a fairly standard way here, many nearby proofs are left out for the purpose of brevity. Instead, we focus on the relationship between the new notions of asymptotic regularity and already available constraint qualifications from nonsmooth optimization. These results are applied to MPCCs in Section 4. As we will see, we precisely recover already available theory from the literature which has been obtained using a standard local decomposition approach. In Section 5, bilevel optimization problems of a special structure are studied in the light of asymptotic stationarity and regularity. Particularly, we state a stationarity condition which holds at all local minimizers, formulate a numerical method which is capable of finding stationary points in this sense, and present some associated numerical results. The paper closes with some concluding remarks in Section 6.
2. Notation and preliminaries

The general notation in this paper is standard. We use $\bar{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$ in order to denote the extended real line. The space $\mathbb{R}^n$ is equipped with the Euclidean norm $\| \cdot \|$. For $z \in \mathbb{R}^n$ and $\varepsilon > 0$, $B_\varepsilon(z) := \{ y \in \mathbb{R}^n \mid \|y - z\| \leq \varepsilon \}$ represents the closed $\varepsilon$-ball around $z$. The distance function $\text{dist}_K : \mathbb{R}^n \to \mathbb{R}$ of a closed, convex set $K \subset \mathbb{R}^n$ is given by $\text{dist}_K(z) := \inf\{\|y - z\| \mid y \in K\}$ for each $z \in \mathbb{R}^n$. Moreover, $\Pi_K : \mathbb{R}^n \to \mathbb{R}^n$ denotes the projection map associated with $K$. Whenever $\phi : \mathbb{R}^n \to \mathbb{R}$ is smooth at some point $\bar{z} \in \mathbb{R}^n$, $\nabla \phi(\bar{z}) \in \mathbb{R}^n$ is used to denote the gradient of $\phi$ at $\bar{z}$. For a function $\Phi : \mathbb{R}^n \to \mathbb{R}^m$ and a vector $y \in \mathbb{R}^m$, the mapping $\langle y, \Phi \rangle : \mathbb{R}^n \to \mathbb{R}$ is given by $\langle y, \Phi \rangle(z) := y^\top \Phi(z)$ for each $z \in \mathbb{R}^n$.

For brevity of notation, a tuple $(z_1, \ldots, z_n)$ of real numbers $z_1, \ldots, z_n \in \mathbb{R}$ will be identified with a vector from $\mathbb{R}^n$ which possesses the components $z_1, \ldots, z_n$. For finite index sets $I_1$ and $I_2$ as well as families $(a_i)_{i \in I_1}, (b_i)_{i \in I_2} \subset \mathbb{R}^n$, we call the pair of families $((a_i)_{i \in I_1}, (b_i)_{i \in I_2})$ positive linearly dependent whenever there exist scalars $\alpha_i \geq 0$ ($i \in I_1$) and $\beta_i$ ($i \in I_2$), not all vanishing simultaneously, such that $\sum_{i \in I_1} \alpha_i a_i + \sum_{i \in I_2} \beta_i b_i = 0$. For a set $A \subset \mathbb{R}^n$ and some point $z \in \mathbb{R}^n$, we use $A + z := \{ a + z \mid a \in A \} =: z + A$ for brevity. Finally, for a set-valued mapping $\Gamma : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ and some point $\bar{z} \in \mathbb{R}^n$, we use

$$\limsup_{z \to \bar{z}} \Gamma(z) := \left\{ \xi \in \mathbb{R}^m \mid \exists \{z^k\}_{k \in \mathbb{N}} \subset \mathbb{R}^n, \exists \{\xi^k\}_{k \in \mathbb{N}} \subset \mathbb{R}^m : z^k \to \bar{z}, \xi^k \to \xi, \xi^k \in \Gamma(z^k) \forall k \in \mathbb{N} \right\}$$

in order to denote the outer (or Painlevé–Kuratowski) limit of $\Gamma$ at $\bar{z}$. Note that $\Gamma(\bar{z}) \subset \limsup_{z \to \bar{z}} \Gamma(z)$ holds always true and that the outer limit is always closed. In a case where $\limsup_{z \to \bar{z}} \Gamma(z) \subset \Gamma(\bar{z})$ is valid, $\bar{z}$ is said to be outer semicontinuous at $\bar{z}$.

2.1. Variational analysis

Subsequently, we recall some notions from nonsmooth analysis and generalized differentiation which can be found, e.g., in [14,15,30].

For a closed set $A \subset \mathbb{R}^n$ and some point $\bar{z} \in A$, the regular and limiting normal cone to $A$ at $\bar{z}$ are given, respectively, by means of

$$\hat{N}_A(\bar{z}) := \{ \xi \in \mathbb{R}^n \mid \forall z \in A : \xi^\top (z - \bar{z}) \leq o(\|z - \bar{z}\|) \},$$

$$N_A(\bar{z}) := \limsup_{z \to \bar{z}, z \in A} \hat{N}_A(z).$$

We note that in situations where $A$ is convex, these cones coincide with the standard normal cone from convex analysis, i.e., we find

$$\hat{N}_A(\bar{z}) = N_A(\bar{z}) = \{ \xi \in \mathbb{R}^n \mid \forall z \in A : \xi^\top (z - \bar{z}) \leq 0 \}$$

in this situation.
For a merely lower semicontinuous function \( \psi : \mathbb{R}^n \to \mathbb{R} \), \( \text{dom } \psi := \{ z \in \mathbb{R}^n \mid |\psi(z)| < \infty \} \) and \( \text{epi } \psi := \{ (z, \alpha) \in \mathbb{R}^n \times \mathbb{R} \mid \psi(z) \leq \alpha \} \) denote its domain and epigraph, respectively. Note that the set \( \text{epi } \psi \) is closed. Let us fix a point \( \bar{z} \in \text{dom } \psi \) where \( \psi \) is locally Lipschitz continuous. Then the set
\[
\widehat{\partial} \psi(\bar{z}) := \{ \xi \in \mathbb{R}^n \mid (\xi, -1) \in \mathcal{N}_{\text{epi } \psi}(\bar{z}, \psi(\bar{z})) \}
\]
is called the regular (or Fréchet) subdifferential of \( \psi \) at \( \bar{z} \). Furthermore, we refer to
\[
\partial \psi(\bar{z}) := \{ \xi \in \mathbb{R}^n \mid (\xi, -1) \in \mathcal{N}_{\text{epi } \psi}(\bar{z}, \psi(\bar{z})) \}
\]
as the limiting (or Mordukhovich) subdifferential of \( \psi \) at \( \bar{z} \). Finally,
\[
\partial^c \psi(\bar{z}) := \text{conv } \partial \psi(\bar{z}),
\]
i.e. the convex hull of \( \partial \psi(\bar{z}) \), is referred to as the Clarke (or convexified) subdifferential of \( \psi \) at \( \bar{z} \). By definition, we have \( \widehat{\partial} \psi(\bar{z}) \subset \partial \psi(\bar{z}) \subset \partial^c \psi(\bar{z}) \), and whenever \( \psi \) is convex, all these sets coincide with the subdifferential of \( \psi \) in the sense of convex analysis. We note that the regular and limiting subdifferential are positive homogeneous while Clarke’s subdifferential is homogeneous.

Some properties of Mordukhovich’s and Clarke’s subdifferential, which we will point out below, are of essential importance in this paper. Therefore, let us assume again that \( \psi \) is locally Lipschitz continuous at \( \bar{z} \). Then \( \partial^\square \psi(\bar{z}) \) is nonempty where \( \partial^\square \) is a representative of the operators \( \partial \) and \( \partial^c \) (here and in the course of this paper). Furthermore, the set-valued mapping \( z \mapsto \partial^\square \psi(z) \) is locally bounded at \( \bar{z} \), i.e. there are a neighbourhood \( U \) of \( \bar{z} \) and a bounded set \( B \subset \mathbb{R}^n \) such that \( \partial^\square \psi(z) \subset B \) is valid for all \( z \in U \). Additionally, the set-valued mapping \( z \mapsto \partial^\square \psi(z) \) is outer semicontinuous at \( \bar{z} \). This property may be also referred to as robustness of the subdifferential \( \partial^\square \).

Below, we present a calculus rule for the subdifferential of minimum functions. Note that we do not only provide upper estimates but precise formulas here.

**Lemma 2.1:** Let \( f_1, f_2 : \mathbb{R}^n \to \mathbb{R} \) be continuously differentiable functions and consider the function \( \varphi : \mathbb{R}^n \to \mathbb{R} \) given by \( \varphi(z) := \min(f_1(z), f_2(z)) \) for all \( z \in \mathbb{R}^n \). For each point \( \bar{z} \in \mathbb{R}^n \), the following formulas hold:

\[
\partial \varphi(\bar{z}) = \begin{cases} 
\{ \nabla f_1(\bar{z}) \} & f_1(\bar{z}) < f_2(\bar{z}), \\
\{ \nabla f_1(\bar{z}), \nabla f_2(\bar{z}) \} & f_1(\bar{z}) = f_2(\bar{z}), \\
\{ \nabla f_2(\bar{z}) \} & f_1(\bar{z}) > f_2(\bar{z}),
\end{cases}
\]

\[
\partial^c \varphi(\bar{z}) = \begin{cases} 
\{ \nabla f_1(\bar{z}) \} & f_1(\bar{z}) < f_2(\bar{z}), \\
\text{conv}\{ \nabla f_1(\bar{z}), \nabla f_2(\bar{z}) \} & f_1(\bar{z}) = f_2(\bar{z}), \\
\{ \nabla f_2(\bar{z}) \} & f_1(\bar{z}) > f_2(\bar{z}),
\end{cases}
\]
Let a distinction of cases. If \([15, \text{Proposition 1.113}]\). In order to validate the converse inclusion, we employ Proposition 2.3.12 and [15, Theorem 3.46(ii)], respectively, while observing that \(f_1 \) and \(f_2 \) are continuously differentiable. Thus, the formula for \(\partial \varphi(\bar{z})\) is obtained from the homogeneity of Clarke’s subdifferential.

It remains to prove the formula for \(\partial \varphi(\bar{z})\). The inclusion \(\subset\) is shown in [15, Proposition 1.113]. In order to validate the converse inclusion, we employ a distinction of cases. If \(f_1(\bar{z}) < f_2(\bar{z})\) holds, then we find \(\varphi(z) = f_1(z)\) locally around \(\bar{z}\), and by continuous differentiability of \(f_1\), \(\partial \varphi(\bar{z}) = \{\nabla f_1(\bar{z})\}\) follows. Similarly, we can address the situation \(f_1(\bar{z}) > f_2(\bar{z})\). Finally, let us investigate the case \(f_1(\bar{z}) = f_2(\bar{z})\). We will show \(\nabla f_1(\bar{z}) \in \partial \varphi(\bar{z})\). Similarly, one obtains \(\nabla f_2(\bar{z}) \in \partial \varphi(\bar{z})\). Suppose that there is a sequence \(\{z^k\}_{k \in \mathbb{N}} \subset \mathbb{R}^n\) such that \(z^k \to \bar{z}\) and \(f_1(z^k) < f_2(z^k)\) for all \(k \in \mathbb{N}\). By continuity of \(f_1\) and \(f_2\) as well as continuous differentiability of \(f_1\), we find \(\hat{\partial} \varphi(z^k) = \{\nabla f_1(z^k)\}\), and taking the limit \(k \to \infty\) yields \(\nabla f_1(\bar{z}) \in \partial \varphi(\bar{z})\). If such a sequence \(\{z^k\}_{k \in \mathbb{N}}\) does not exist, we find a neighbourhood \(U\) of \(\bar{z}\) such that \(f_1(z) \geq f_2(z)\) holds for all \(z \in U\). On the one hand, this shows \(\varphi(z) = f_2(z)\) for all \(z \in U\) and, thus, \(\partial \varphi(\bar{z}) = \{\nabla f_2(\bar{z})\}\). On the other hand, due to \(f_1(\bar{z}) = f_2(\bar{z})\), \(\bar{z}\) is a local minimizer of \(f_1 - f_2\) which yields \(\nabla f_1(\bar{z}) - \nabla f_2(\bar{z}) = 0\). Summing this up, we have shown \(\nabla f_1(\bar{z}) \in \partial \varphi(\bar{z})\).

\[\partial(-\varphi)(\bar{z}) = \partial^c(-\varphi)(\bar{z}) = \begin{cases} \{-\nabla f_1(\bar{z})\} & f_1(\bar{z}) < f_2(\bar{z}), \\ \text{conv}\{-\nabla f_1(\bar{z}), -\nabla f_2(\bar{z})\} & f_1(\bar{z}) = f_2(\bar{z}), \\ \{-\nabla f_2(\bar{z})\} & f_1(\bar{z}) > f_2(\bar{z}). \end{cases}\]

**Proof:** Exploiting \(-\varphi(z) = \max(-f_1(z), -f_2(z))\) which holds for each \(z \in \mathbb{R}^n\), the formulas for \(\partial^c(-\varphi)(\bar{z})\) and \(\partial(-\varphi)(\bar{z})\) follow from the maximum rules [14, Proposition 2.3.12] and [15, Theorem 3.46(ii)], respectively, while observing that \(f_1 \) and \(f_2 \) are continuously differentiable. Thus, the formula for \(\partial \varphi(\bar{z})\) is obtained from the homogeneity of Clarke’s subdifferential.

### 2.2. Sequential stationarity for optimization problems with Lipschitzian geometric constraints

In this section, we investigate the optimization problem

\[\min\{f(z) \mid F(z) \in K\}, \quad (Q)\]

where \(f: \mathbb{R}^n \to \overline{\mathbb{R}}\) and \(F: \mathbb{R}^n \to \overline{\mathbb{R}}^m\) are given functions and \(K \subset \mathbb{R}^m\) is a convex polyhedral set. Let \(\mathcal{F} := \{z \in \mathbb{R}^n \mid F(z) \in K\}\) be the feasible set of \((Q)\). For later use, we are going to characterize local minimizers of \((Q)\) with the aid of sequential stationarity conditions which are based on the limiting subdifferential. Related results can be found in [7,13].

**Proposition 2.2:** Let \(\tilde{z} \in \mathcal{F}\) be a local minimizer of \((Q)\). Furthermore, assume that \(f\) and \(F\) are Lipschitz continuous around \(\tilde{z}\). Then there exist sequences \(\{z^k\}_{k \in \mathbb{N}}, \{e^k\}_{k \in \mathbb{N}} \subset \mathbb{R}^n\) and \(\{\lambda^k\}_{k \in \mathbb{N}} \subset \mathbb{R}^m\) such that \(z^k \to \tilde{z}\), \(e^k \to 0\), and

\[\forall k \in \mathbb{N}: \; e^k \in \partial f(z^k) + \partial \langle \lambda^k, F(z^k) \rangle, \quad (1a)\]
Since \( f \) feasible set, this program possesses a global minimizer \( z_k \). Noting that the objective function of \((Q(z_k))\) guarantees 0 there is some \( f \) and thus, we have \( \bar{z} \). Additionally, we may assume w.l.o.g. that there is some \( z \in \mathbb{B}_\delta(\tilde{z}) \) such that \( z_k \to \tilde{z} \) holds as \( k \to \infty \). By feasibility of \( \bar{z} \), we infer

\[
\forall k \in \mathbb{N} : f(z_k) + \frac{k}{2} \text{dist}_K^2(F(z_k)) + \frac{1}{2} \|z_k - \bar{z}\|^2 \leq f(\bar{z}).
\]

Since \( f \) is continuous on \( \mathbb{B}_\delta(\bar{z}) \), \( \{f(z_k)\}_{k \in \mathbb{N}} \) is a bounded sequence. Thus, we find a constant \( C > 0 \) which satisfies \( \frac{k}{2} \text{dist}_K^2(F(z_k)) \leq C \). Consequently, we have \( \text{dist}_K^2(F(z_k)) \to 0 \) as \( k \to \infty \). Exploiting the continuity of the distance function as well as \( F \), this yields \( F(\bar{z}) \in K \), i.e. \( \bar{z} \in \mathcal{F} \cap \mathbb{B}_\delta(\tilde{z}) \). By choice of \( \delta \), this leads to

\[
f(\bar{z}) \leq f(\bar{z}) + \frac{1}{2} \|\bar{z} - \tilde{z}\|^2
\]

\[
\leq \limsup_{k \to \infty} \left( f(z_k) + \frac{k}{2} \text{dist}_K^2(F(z_k)) + \frac{1}{2} \|z_k - \tilde{z}\|^2 \right)
\]

\[
\leq f(\bar{z}) \leq f(\tilde{z}),
\]

and thus, we have \( \bar{z} = \tilde{z} \). Additionally, we may assume w.l.o.g. that \( \{z_k\}_{k \in \mathbb{N}} \) belongs to the interior of \( \mathbb{B}_\delta(\tilde{z}) \).

Next, we define functions \( f_1, f_2 : \mathbb{R}^n \to \mathbb{R} \) by means of

\[
\forall z \in \mathbb{R}^n : f_1(z) := \frac{1}{2} \text{dist}_K^2(F(z)), \quad f_2(z) := \frac{1}{2} \|z - \tilde{z}\|^2.
\]

Recalling that \( z_k \) is a minimizer of \((Q(z_k))\), Fermat’s rule (see [15, Proposition 1.114]) guarantees 0 belongs to the interior of \( \mathbb{B}_\delta(\tilde{z}) \) for each \( k \in \mathbb{N} \). Since \( f_2 \) is continuously differentiable with gradient \( \nabla f_2(z) = z - \tilde{z} \) for arbitrary \( z \in \mathbb{R}^n \), \( \tilde{z} - z_k \in \widehat{\partial}(f + k f_1 + f_2)(z_k) \) follows from [15, Proposition 1.107]. Next, we apply the sum rule for the limiting subdifferential (see [15, Theorem 3.36]) in order to find that \( \tilde{z} - z_k \in \partial f(z_k) + k \partial f_1(z_k) \) holds for each \( k \in \mathbb{N} \). Due to \( f_1 = \frac{1}{2} \text{dist}_K^2 \circ F \), continuous differentiability of the squared distance function to a convex set, and
Lipschitzianity of $F$ at $z^k$, we find

$$\partial f_1(z^k) = \partial (F(z^k) - \Pi_K(F(z^k)), F(z^k))$$

from the subdifferential chain rule [15, Corollary 3.43]. Setting $\varepsilon^k := \bar{z} - z^k$ and $\lambda^k := k(F(z^k) - \Pi_K(F(z^k)))$ for each $k \in \mathbb{N}$, we find $\varepsilon^k \to 0$ and (1a). By construction, we have $z^k \to \bar{z}$ and $\lambda^k \in \hat{N}_K(F(z^k))$ for each $k \in \mathbb{N}$ (see [30, Section 6.E]). Due to $F(z^k) \to F(\bar{z})$ and the polyhedrality of $K$, we can apply [31, Lemma 2.1] in order to find (1b) for large enough $k \in \mathbb{N}$. This completes the proof.

Let us comment on the assertion of Proposition 2.2. First, we would like to point the reader’s attention to the fact that the appearing multiplier sequence $\{\lambda^k\}_{k \in \mathbb{N}}$ does not need to be bounded. If this would be the case, then one could simply take the limit along some convergent subsequence in (1) in order to find some $\lambda \in \mathbb{R}^m$ which satisfies

$$0 \in \partial f(\bar{z}) + \partial \langle \lambda, F(\bar{z}) \rangle, \quad \lambda \in \hat{N}_K(F(\bar{z})).$$

This follows by the robustness of the limiting subdifferential (see [13, Lemma 3.4] as well) and convexity of $K$. Note that due to the appearance of the limiting subdifferential, these conditions precisely correspond to the so-called Mordukhovich (or simply M-) stationarity conditions of (Q) at $\bar{z}$ (recall that since $K$ is convex, the limiting and regular normal cone to this set coincide). Clearly, there exist optimization problems whose local minimizers are not M-stationary so that it is completely reasonable that we are not in a position to show boundedness of $\{\lambda^k\}_{k \in \mathbb{N}}$ without additional regularity in the proof of Proposition 2.2.

Next, we would like to point out that the conditions in Proposition 2.2 precisely correspond to the so-called AM-stationarity conditions of (Q), which were introduced in [13, Definition 3.1] for much more general problems. This can be seen by employing the scalarization property of the so-called limiting coderivative (see [15, Theorem 1.90]).

Finally, let us mention that it is also possible to state the assertion of Proposition 2.2 in terms of the regular subdifferential. However, one has to exploit the so-called fuzzy sum rule during the proof (see [15, Theorem 2.33]) since the regular subdifferential does not obey a classical sum rule. Respecting this, one would have to replace (1a) by

$$\forall k \in \mathbb{N}: \varepsilon^k \in \hat{\partial}f(z^k_0) + \hat{\partial} \langle \lambda^k, F(z^k) \rangle,$$

where $\{z^k_0\}_{k \in \mathbb{N}}, \{z^k_C\}_{k \in \mathbb{N}} \subset \mathbb{R}^n$ are sequences satisfying $z^k_0 \to \bar{z}$ and $z^k_C \to \bar{z}$. By definition of the limiting subdifferential, one can show that, in general, this condition is not stronger than the one postulated in Proposition 2.2 as long as both $f$ and $F$ are nonsmooth around $\bar{z}$. Let us also note that when taking the limit in (2),
one would end up with a condition in terms of the limiting subdifferential anyway. That is why we rely on the statement of Proposition 2.2 in the remainder of the paper.

3. Asymptotic stationarity and regularity for Lipschitzian nonlinear programs

In the past, several tools of generalized differentiation have been introduced which allow transferring the Karush–Kuhn–Tucker (KKT) theory for standard nonlinear programs with continuously differentiable data functions to a non-smooth framework. Amongst others, let us mention the subdifferential constructions introduced by Clarke and Mordukhovich (see [14,15]) which enjoy (almost) full calculus and can be used to derive KKT-type necessary optimality conditions for the model problem (P) (see [[15, Section 5.1.3];[32, Section 5.6]]). We will refer to these conditions as the systems of $\partial^c$- and $\partial$-stationarity, respectively. In the literature, the nomenclatures of Clarke (or simply C-) and M-stationarity are also common, but we will avoid these terms here for some reasons which will become clear in the course of the paper. For the purpose of completeness, we start our investigations by stating a precise definition of $\partial^c$- and $\partial$-stationarity, respectively.

Throughout the section, let $Z$ denote the feasible set of (P). We implicitly assume that whenever $\bar{z} \in Z$ is a fixed feasible point of (P), then the functions $\varphi_0, \ldots, \varphi_{p+q}$ are locally Lipschitz continuous in a neighbourhood of $\bar{z}$. Furthermore, we make use of the so-called index set associated with inequality constraints active at $\bar{z}$ which is given by $I(\bar{z}) := \{i \in I | \varphi_i(\bar{z}) = 0\}$. Finally, recall that $\partial$ plays the role of the subdifferential operator $\partial^c$ or $\partial$.

**Definition 3.1:** A feasible point $\bar{z} \in Z$ of (P) is called $\partial$-stationary whenever there are multipliers $\lambda \in \mathbb{R}^{p+q}$ which satisfy the following conditions:

\[
0 \in \partial \varphi_0(\bar{z}) + \sum_{i \in I} \lambda_i \partial \varphi_i(\bar{z}) + \sum_{i \in J} \lambda_i \left( \partial \varphi_i(\bar{z}) \cup \partial (-\varphi_i)(\bar{z}) \right),
\]

\[
\forall i \in I: \quad \min(\lambda_i, -\varphi_i(\bar{z})) = 0,
\]

\[
\forall i \in J: \quad \lambda_i \geq 0.
\]

Although $\partial$-stationarity is sharper than $\partial^c$-stationarity, it might be beneficial to work with Clarke’s subdifferential in some situations since it is far easier to compute (see [14, Theorem 2.5.1]) and its homogeneity allows for an easier calculus (see, e.g. Section 5 where this issue is of essential importance). Similar arguments justify the consideration of constraint qualifications based on Clarke’s subdifferential. Let us point out that in both stationarity systems, the limiting subdifferential is used as the generalized derivative for the objective function. At the first glance, this seems to be uncommon. However, since most of the variational
issues one has to face during the theoretical treatment of \((P)\) are related to the structure of the feasible set (see, e.g. Sections 4 and 5), this choice is reasonable and induces a system of \(\partial^c\)-stationarity which is slightly sharper than the classical one from [32].

Let us put the stationarity concepts from Definition 3.1 into some context. For that purpose, let \(\Phi : \mathbb{R}^n \to \mathbb{R}^{p+q}\) be the vector function whose components are precisely \(\varphi_1, \ldots, \varphi_{p+q}\). We note that \((P)\) is a particular instance of the problem \((Q)\) discussed in Section 2.2 where we fix \(f := \varphi_0, F := \Phi,\) and

\[ K := \{ y \in \mathbb{R}^{p+q} | \forall i \in I: y_i \leq 0, \forall i \in J: y_i = 0 \}. \]

A simple evaluation of the regular normal cone to this particular set \(K\) reveals that

\[ 0 \in \partial \varphi_0(\bar{z}) + \partial \langle \tilde{\lambda}, \Phi \rangle(\bar{z}), \] (4a)

\[ \forall i \in I: \min(\tilde{\lambda}_i, -\varphi_i(\bar{z})) = 0 \] (4b)

for some multiplier \(\tilde{\lambda} \in \mathbb{R}^{p+q}\) might be a reasonable candidate for the \(\partial^\square\)-stationarity system as well. The sum rule for Clarke’s subdifferential as well as its homogeneity (see [14, Section 2.3]) imply

\[ \partial^c(\tilde{\lambda}, \Phi)(\bar{z}) \subset \sum_{i \in I \cup J} \tilde{\lambda}_i \partial \varphi_i(\bar{z}), \]

and due to

\[ \gamma \partial \psi(z) = \begin{cases} |\gamma| \partial^c \psi(z) & \gamma \geq 0, \\ |\gamma| \partial^c(-\psi)(z) & \gamma < 0, \end{cases} \]

for each function \(\psi : \mathbb{R}^n \to \overline{\mathbb{R}}\) which is locally Lipschitzian around \(z \in \mathbb{R}^n\) and each constant \(\gamma \in \mathbb{R}\), the stationarity condition (4) for Clarke’s subdifferential is slightly stronger than \(\partial^c\)-stationarity from Definition 3.1. On the other hand, we find the inclusion

\[ \partial(\tilde{\lambda}, \Phi)(\bar{z}) \subset \sum_{i \in I} \tilde{\lambda}_i \partial \varphi_i(\bar{z}) + \sum_{i \in J} |\tilde{\lambda}_i| (\partial \varphi_i(\bar{z}) \cup \partial (-\varphi_i)(\bar{z})) \]

by the sum rule for the limiting subdifferential (see [15, Theorem 3.36]) and its positive homogeneity. Thus, (3a) from Definition 3.1 for the limiting subdifferential might be weaker than condition (4a) from above. However, system (3) is stated in fully explicit way w.r.t. the subdifferentials of the appearing constraint functions. It is, thus, reasonable to work with the generalized stationarity notions from Definition 3.1 and not with the potentially sharper conditions from (4).

Recently, the concept of \textit{asymptotic} stationarity has attracted lots of attention due to two basic observations. First, some algorithms from optimization theory naturally produce a sequence of iterates whose accumulation points satisfy
such asymptotic stationarity conditions. Second, asymptotic stationarity gives rise to the definition of very weak constraint qualifications. We refer the interested reader to [1–4,6,7,11,13] and the references therein for validation. Below, we present two natural extensions of asymptotic stationarity which apply to the Lipschitzian optimization problem (P) and are based on \( \partial^c \) - and \( \partial \)-stationarity from Definition 3.1.

Definition 3.2: A feasible point \( \bar{z} \in Z \) of (P) is called asymptotically \( \partial^\square \)-stationary (A\( \partial^\square \)-stationary for short) whenever there are sequences \( \{ z^k \}_{k \in \mathbb{N}} \), \( \{ \epsilon^k \}_{k \in \mathbb{N}} \subset \mathbb{R}^n \) and \( \{ \lambda^k \}_{k \in \mathbb{N}} \subset \mathbb{R}^{p+q} \) which satisfy

\[
\epsilon^k \in \partial \phi_0(z^k) + \sum_{i \in I} \lambda^k_i \partial \varphi_i(z^k) + \sum_{i \in J} \lambda^k_i (\partial \varphi_i(z^k) \cup \partial \varphi_i(-\varphi_i(z^k))),
\]

\[
\forall i \in I: \quad \min(\lambda^k_i, -\varphi_i(\bar{z})) = 0,
\]

\[
\forall i \in J: \quad \lambda^k_i \geq 0
\]

for all \( k \in \mathbb{N} \) as well as \( z^k \to \bar{z} \) and \( \epsilon^k \to 0 \).

By definition, each A\( \partial^\square \)-stationary point is A\( \partial^c \)-stationary, but the converse statement does not hold true in general (see Example 3.9). Referring to the considerations at the beginning of this section, we would like to note that due to homogeneity of Clarke's subdifferential, \( \bar{z} \in \mathbb{R}^n \) is A\( \partial^c \)-stationary for (P) if and only if there exist sequences \( \{ z^k \}_{k \in \mathbb{N}} \), \( \{ \epsilon^k \}_{k \in \mathbb{N}} \subset \mathbb{R}^n \) and \( \{ \tilde{\lambda}^k \}_{k \in \mathbb{N}} \subset \mathbb{R}^{p+q} \) which satisfy \( z^k \to \bar{z} \), \( \epsilon^k \to 0 \), as well as

\[
\forall k \in \mathbb{N}: \quad \epsilon^k \in \partial \phi_0(z^k) + \sum_{i \in I \cup J} \tilde{\lambda}^k_i \partial^c \varphi_i(z^k), \quad \min(\tilde{\lambda}^k_i, -\varphi_i(\bar{z})) = 0 \quad (i \in I).
\]

Particularly, the sign condition (5c) on the multipliers associated with equality constraints needs to be dropped in this form of the definition.

In the lemma below, we present an equivalent definition of A\( \partial^\square \)-stationarity which might be more convenient in the light of algorithmic applications since it allows for certain violations of (5b) and (5c).

Lemma 3.3: A feasible point \( \bar{z} \in Z \) of (P) is A\( \partial^\square \)-stationary if and only if there are sequences \( \{ z^k \}_{k \in \mathbb{N}} \), \( \{ \epsilon^k \}_{k \in \mathbb{N}} \subset \mathbb{R}^n \) and \( \{ \lambda^k \}_{k \in \mathbb{N}} \subset \mathbb{R}^{p+q} \) which satisfy \( z^k \to \bar{z} \), \( \epsilon^k \to 0 \), (5a) for each \( k \in \mathbb{N} \), as well as

\[
\forall i \in I: \quad \lim_{k \to \infty} \min \lim_{k \to \infty} (\lambda^k_i, -\varphi_i(z^k)) = 0,
\]

\[
\forall i \in J: \quad \liminf_{k \to \infty} \lambda^k_i \geq 0.
\]

Proof: [\( \Rightarrow \]): If \( \bar{z} \) is A\( \partial^\square \)-stationary, we find sequences \( \{ z^k \}_{k \in \mathbb{N}} \), \( \{ \epsilon^k \}_{k \in \mathbb{N}} \subset \mathbb{R}^n \) and \( \{ \lambda^k \}_{k \in \mathbb{N}} \subset \mathbb{R}^{p+q} \) satisfying (5) for each \( k \in \mathbb{N} \), \( z^k \to \bar{z} \), and \( \epsilon^k \to 0 \). This
already yields (6b). By continuity of \( \varphi_i \) at \( \bar{z} \), we find \( \varphi_i(z^k) \to \varphi_i(\bar{z}) \leq 0 \) for each \( i \in I \). Furthermore, (5b) guarantees \( \lambda^k_i \geq 0 \) for all \( k \in \mathbb{N} \) and \( i \in I \). For \( i \in I \setminus I(\bar{z}) \), we have \( \lambda^k_i = 0 \) and \( \varphi_i(z^k) < 0 \) for all sufficiently large \( k \in \mathbb{N} \) which yields \( \min(\lambda^k_i, -\varphi_i(z^k)) = 0 \) for large enough \( k \in \mathbb{N} \). Fixing \( i \in I(\bar{z}) \), we find \( \varphi_i(z^k) \to 0 \) which yields \( \min(\lambda^k_i, -\varphi_i(z^k)) \to 0 \). This shows validity of (6a).

\[ \leftarrow : \] Assume that there are sequences \( (z^k)_{k \in \mathbb{N}}, (e^k)_{k \in \mathbb{N}} \subseteq \mathbb{R}^n \) and \( (\lambda^k_i)_{k \in \mathbb{N}} \subseteq \mathbb{R}^{p+q} \) which satisfy \( z^k \to \bar{z}, e^k \to 0 \), (5a) for each \( k \in \mathbb{N} \), and (6). Thus, for each \( i \in I \), we find a sequence \( \{\xi^k_i\}_{k \in \mathbb{N}} \subseteq \mathbb{R}^n \) with \( \xi^k_i \in \partial \varphi_i(z^k) \) for each \( k \in \mathbb{N} \), and for each \( i \in J \), we find a sequence \( \{\eta^k_i\}_{k \in \mathbb{N}} \subseteq \mathbb{R}^n \) with \( \eta^k_i \in \partial \varphi_i(z^k) \). Then (6b) guarantees that \( \min(\lambda^k_i, -\varphi_i(z^k)) = 0 \) for each \( i \in I \). Performing the above transformations iteratively for each index \( j \in I \cup J \) where a violation of (5b) or (5c) occurs, we can hide these asymptotic violations,
restricted via (6a) and (6b), in the definition of \( \{ \varepsilon_k \}_{k \in \mathbb{N}} \). Hence, \( \bar{z} \) is A\( \partial \square \)-stationary.

Let us now invoke Proposition 2.2. Exploiting the sum rule and positive homogeneity of the limiting subdifferential, the following result follows easily by similar considerations as presented after Definition 3.1.

**Theorem 3.4:** If \( \bar{z} \in \mathcal{Z} \) is a local minimizer of (P), then it is an A\( \partial \square \)-stationary point of this program.

From the above result, we immediately see that each local minimizer of (P) is A\( \partial c \)-stationary as well. Using Clarke’s subdifferential, approximate KKT-type necessary optimality conditions can be found in [33]. Furthermore, we would like to mention the recently published paper [12] where Goldstein’s \( \varepsilon \)-subdifferential construction is used to design a sequential stationarity condition for Lipschitzian programs. The authors stated an implementable algorithm which computes asymptotically stationary points in their sense. On the other hand, Goldstein’s \( \varepsilon \)-subdifferential is even larger than Clarke’s subdifferential (w.r.t. set inclusion) and, thus, provides very weak stationarity conditions. Furthermore, its numerical computation is quite challenging since it is likely to be set-valued for each point from the underlying function’s domain.

Next, we present a simple observation regarding the sequential stationarity notions from Definition 3.2. Its proof is based on the outer semicontinuity of the Clarke and limiting subdifferential as well as their local boundedness for locally Lipschitzian functions and, for large parts, can be distilled from the proof of [13, Lemma 3.4]. It basically says that each asymptotically \( \partial \square \)-stationary point of (P) satisfies a Fritz–John–type condition based on the subdifferential construction \( \partial \square \). Simple examples indicate, however, that A\( \partial \square \)-stationarity is, in general, stronger than these Fritz–John-type conditions.

**Lemma 3.5:** Let \( \bar{z} \in \mathcal{Z} \) be an A\( \partial \square \)-stationary point of (P) such that the sequences \( \{ z^k \}_{k \in \mathbb{N}} \), \( \{ \varepsilon^k \}_{k \in \mathbb{N}} \subset \mathbb{R}^n \) and \( \{ \lambda^k \}_{k \in \mathbb{N}} \subset \mathbb{R}^{p+q} \) with \( z^k \to \bar{z} \) and \( \varepsilon^k \to 0 \) satisfy (5) for each \( k \in \mathbb{N} \). Then the following assertions hold.

(a) If \( \{ \lambda^k \}_{k \in \mathbb{N}} \) is bounded, then \( \bar{z} \) is \( \partial \square \)-stationary.
(b) If \( \{ \lambda^k \}_{k \in \mathbb{N}} \) is not bounded, then we find a nonzero vector \( \lambda \in \mathbb{R}^{p+q} \) which satisfies (3b), (3c), and

\[
0 \in \sum_{i \in I} \lambda_i \partial \square \varphi_i(\bar{z}) + \sum_{i \in J} \lambda_i (\partial \square \varphi_i(\bar{z}) \cup \partial \square (-\varphi_i)(\bar{z})).
\]

Below, we are going to interrelate \( \partial \square \)- and A\( \partial \square \)-stationary points of (P). Therefore, we fix a feasible point \( \bar{z} \in \mathcal{Z} \) of (P). Let us introduce a set-valued mapping
\[ \overline{\mathcal{M}} : \mathbb{R}^n \rightarrow \mathbb{R}^n \] by means of
\[ \overline{\mathcal{M}}(z) := \left\{ \sum_{i \in I(\bar{z})} \lambda_i \partial \Box \varphi_i(z) + \sum_{i \in J} \lambda_i \left( \partial \Box \varphi_i(z) \cup \partial \Box (-\varphi_i)(z) \right) \mid \lambda_i \geq 0 \right\} \]
for each \( z \in \mathbb{R}^n \). Note that this map explicitly depends on \( \bar{z} \) since the set \( I(\bar{z}) \) appears. The definition of \( \overline{\mathcal{M}} \) directly shows that \( \bar{z} \) is \( \partial \Box \)-stationary for (P) if and only if \( \partial \varphi_0(\bar{z}) \cap (-\overline{\mathcal{M}}(\bar{z})) = \emptyset \) holds. Furthermore, by definition of \( A \partial \Box \)-stationarity, we find the following result. Its proof is analogous to the one of [13, Lemma 3.6] and, again, basically exploits outer semicontinuity and local boundedness of the respective subdifferential as well as the definition of the outer set limit.

**Lemma 3.6:** Let \( \bar{z} \in \mathcal{Z} \) be a feasible point of (P). Then the following assertions hold.

(a) If \( \bar{z} \) is an \( A \partial \Box \)-stationary point of (P), then \( \partial \varphi_0(\bar{z}) \cap (-\limsup_{z \to \bar{z}} \overline{\mathcal{M}}(z)) \neq \emptyset \).

(b) If \( \varphi_0 \) is continuously differentiable at \( \bar{z} \in \mathbb{R}^n \) while \( -\nabla \varphi_0(\bar{z}) \in \limsup_{z \to \bar{z}} \overline{\mathcal{M}}(z) \) is valid, then \( \bar{z} \) is an \( A \partial \Box \)-stationary point of (P).

Taking Theorem 3.4 and Lemma 3.6 together, the definition of the following constraint qualifications is reasonable.

**Definition 3.7:** Let \( \bar{z} \in \mathcal{Z} \) be a feasible point of (P).

(a) We call \( \bar{z} \) asymptotically \( \partial \Box \)-regular (\( A \partial \Box \)-regular for short) whenever the condition \( \limsup_{z \to \bar{z}} \overline{\mathcal{M}}(z) \subset \overline{\mathcal{M}}(\bar{z}) \) is valid, i.e. if \( \overline{\mathcal{M}} \) is outer semicontinuous at \( \bar{z} \).

(b) We call \( \bar{z} \) weakly asymptotically \( \partial \)-regular (\( wA \partial \)-regular for short) whenever the condition \( \limsup_{z \to \bar{z}} \overline{\mathcal{M}}(z) \subset \overline{\mathcal{M}}^c(\bar{z}) \) is valid.

We would like to point out that \( A \partial \Box \)-regularity and \( wA \partial \)-regularity reduce to the so-called cone continuity property from [3, Definition 3.1], which is sometimes referred to as AKKT-regularity, whenever the functions \( \varphi_i \) \( (i \in I(\bar{z}) \cup J) \) are continuously differentiable at \( \bar{z} \). In the general nonsmooth setting, however, we only get the relations

\[ A \partial \text{-regularity} \quad \implies \quad wA \partial \text{-regularity}, \]
\[ A \partial^c \text{-regularity} \quad \implies \quad wA \partial \text{-regularity}. \]
The following examples underline that $A\partial$-regularity and $A\partial^c$-regularity are independent of each other. Note that both of these examples are stated in the context of complementarity constrained optimization (see Section 4 as well). Actually, Example 3.9 is taken from [6, Example 6] where it is used to visualize closely related issues. Finally, let us mention that these examples also indicate that $wA\partial$-regularity is strictly weaker than $\partial$- and $\partial^c$-regularity.

**Example 3.8:** We consider the feasible region $\mathcal{Z} \subset \mathbb{R}^2$ modelled by

$$\mathcal{Z} := \{ z \in \mathbb{R}^2 | \varphi_1(z) := z_1^3 - z_2 \leq 0, \varphi_2(z) := \min(z_1, z_2) = 0 \}$$

at $\bar{z} := (0, 0)$. For the computation of the subdifferentials associated with $\varphi_2$, we refer the reader to Lemma 2.1. Exploiting $z^k := ((3k)^{-1/2}, 0)$, we find

$$(1, 1) = k (1/k, -1) + (k + 1) (0, 1) \in k \nabla \varphi_1(z^k) + (k + 1) \partial \varphi_2(z^k) \subset \overline{\mathcal{M}}(z^k)$$

for each $k \in \mathbb{N}$. Thus, due to $\overline{\mathcal{M}}(\bar{z}) = \{ \eta \in \mathbb{R}^2 | \eta_1 = 0 \lor \eta_2 \leq 0 \}$, $\bar{z}$ is not $A\partial$-regular.

On the other hand, we find $\overline{\mathcal{M}}^c(\bar{z}) = \{ \eta \in \mathbb{R}^2 | \eta_1 \geq 0 \lor \eta_2 \leq 0 \}$. Suppose now that there is some $\eta \in \limsup_{z \to \bar{z}} \overline{\mathcal{M}}^c(z^k)$ which satisfies $\eta_1 < 0$ and $\eta_2 > 0$, i.e. that $\bar{z}$ is not $A\partial^c$-regular. Then we find $\{ z^k \}_{k \in \mathbb{N}}, \{ \eta^k \}_{k \in \mathbb{N}} \subset \mathbb{R}^2$ such that $z^k \to \bar{z}$, $\eta^k \to \eta$, and $\eta^k \in \overline{\mathcal{M}}^c(z^k)$ for all $k \in \mathbb{N}$, i.e. there are sequences $\{ \lambda^k_1 \}_{k \in \mathbb{N}}, \{ \lambda^k_2 \}_{k \in \mathbb{N}} \subset \mathbb{R}_+$ and $\{ \xi^k \}_{k \in \mathbb{N}} \subset \mathbb{R}^2$ such that

$$\forall k \in \mathbb{N}: \eta^k = \lambda^k_1 (3(z^k_1)^2, -1) + \lambda^k_2 \xi^k, \quad \xi^k \in \partial^c \varphi_2(z^k) \cup \partial^c (-\varphi_2)(z^k).$$

Due to $\eta^k_1 < 0$ for sufficiently large $k \in \mathbb{N}$, we find $\xi^k_1 < 0$ for sufficiently large $k \in \mathbb{N}$. This is only possible if $z^k_1 \leq z_2^k$ is valid for sufficiently large $k \in \mathbb{N}$. In these situations, we have $\xi^k_2 \leq 0$ as well, i.e. $\eta^k_2 \leq 0$ follows for large enough $k \in \mathbb{N}$. This, however, contradicts $\eta_2 > 0$. Thus, $\bar{z}$ is $A\partial^c$-regular.

**Example 3.9:** Let us investigate the feasible region $\mathcal{Z} \subset \mathbb{R}^3$ given by

$$\mathcal{Z} := \left\{ z \in \mathbb{R}^3 \left| \begin{array}{l} \varphi_1(z) := -z_1 \leq 0, \varphi_2(z) := -z_3 \leq 0, \\
\varphi_3(z) := \min(z_1^3 + z_2 + z_3, z_1^3 - z_2 + z_3) = 0 \end{array} \right. \right\}$$

at $\bar{z} := (0, 0, 0)$. For the computation of the subdifferentials associated with $\varphi_3$, we make use of Lemma 2.1 again. Considering $z^k := ((3k/2)^{-1/2}, 0, 0)$, we find

$$(1, 0, 0) = (-1, 0, 0) + k (0, 0, -1) + k (2/k, 0, 1)$$

$$\in \nabla \varphi_1(z^k) + k \nabla \varphi_2(z^k) + k \partial \varphi_3(z^k) \subset \overline{\mathcal{M}}^c(z^k)$$

for each $k \in \mathbb{N}$. On the other hand, a simple calculation reveals the relation $\overline{\mathcal{M}}^c(\bar{z}) \subset \mathbb{R}_- \times \mathbb{R}^2$ which means that $\bar{z}$ cannot be $A\partial^c$-regular.
One can show that \( \overline{M}(\bar{z}) = \{ \eta \in \mathbb{R}^3 \mid \eta_1 \leq 0, \eta_3 \leq |\eta_2| \} \) holds. Fix some \( \eta \in \limsup_{z \to \bar{z}} \overline{M}(z) \). Then we find sequences \( \{z^k\}_{k \in \mathbb{N}}, \{\eta^k\}_{k \in \mathbb{N}} \subset \mathbb{R}^3 \) such that \( z^k \to \bar{z}, \eta^k \to \eta \), and \( \eta^k \in \overline{M}(z^k) \) for each \( k \in \mathbb{N} \). One can easily check that

\[
\partial \varphi_3(z) \cup \partial(-\varphi_3)(z) \subset \{(3z_1^2, \pm 1, 1)\} \cup \{(-3z_1^2, \alpha, -1) \mid \alpha \in [-1, 1]\}
\]

holds for all \( z \in \mathbb{R}^3 \), and this reveals that \( \eta_3^k \leq |\eta_2^k| \) is valid for all \( k \in \mathbb{N} \). Taking the limit \( k \to \infty \) yields \( \eta_3 \leq |\eta_2| \). Supposing that \( \eta_1 > 0 \) holds, there are sequences \( \{\lambda_1^k\}_{k \in \mathbb{N}}, \{\lambda_2^k\}_{k \in \mathbb{N}}, \{\lambda_3^k\}_{k \in \mathbb{N}} \subset \mathbb{R}_+ \) such that

\[
\eta^k = \lambda_1^k(-1, 0, 0) + \lambda_2^k(0, 0, -1) + \lambda_3^k(3(z_1^k)^2, \pm 1, 1)
\]

is valid for sufficiently large \( k \in \mathbb{N} \). Inspecting the second component, \( \{\lambda_3^k\}_{k \in \mathbb{N}} \) needs to be convergent since \( \{\eta_2^k\}_{k \in \mathbb{N}} \) converges to \( \eta_2 \). Thus, we find \( \lambda_3^k 3(z_1^k)^2 \to 0 \) by \( z_1^k \to 0 \), and due to \( \eta_1^k \to \eta_1 \), this leads to \( \eta_1 \leq 0 \) – a contradiction. As a consequence, \( \bar{z} \) is \( A\partial\)-regular.

Note that using the objective function given by \( \varphi_0(z) := -z_1 \) for all \( z \in \mathbb{R}^3 \), one can exploit Lemma 3.6 in order to see that \( \bar{z} \) is an \( A\partial^c \)-stationary point of the associated program \( (P) \) which is not \( A\partial \)-stationary.

Exploiting Lemma 3.6, we find that each \( A\partial\Box \)-stationary point of \( (P) \), which is \( A\partial\Box \)-regular, is already \( \partial\Box \)-stationary. More precisely, \( A\partial\Box \)-regularity is the weakest condition which implies that an \( A\partial\Box \)-stationary point is actually \( \partial\Box \)-stationary. In the light of [3, Section 1], we may thus refer to \( A\partial\Box \)-regularity as a strict constraint qualification. Taking Theorem 3.4 and Lemma 3.6 together, we find the following result.

**Theorem 3.10:** Let \( \bar{z} \in \mathcal{Z} \) be a local minimizer of \( (P) \). Then the following assertions hold.

(a) If \( \bar{z} \) is \( wA\partial \)-regular, then \( \bar{z} \) is \( \partial^c \)-stationary.
(b) If \( \bar{z} \) is \( A\partial \)-regular, then \( \bar{z} \) is \( \partial \)-stationary.

Exploiting the sequential stationarity condition based on the regular subdifferential mentioned at the end of Section 2.2, it is technically possible to introduce asymptotic regularity conditions based on the regular subdifferential as well. However, let us note that this approach comes along with two disadvantages. First, the regular subdifferential does not obey a classical but only a fuzzy sum rule which is why the outer semicontinuity properties of much more difficult set-valued mappings would need to be considered. Second, let us recall that in contrast to the limiting and Clarke subdifferential, the regular subdifferential is not outer semicontinuous in general (see Section 2.1) so that a constraint qualification for stationarity in terms of the regular subdifferential is likely to fail anyway.
In the remaining part of this section, we are going to embed the constraint qualifications from Definition 3.7 into the landscape of qualification conditions from nonsmooth optimization. Let us recall that \(\partial\)-NMFCQ, the nonsmooth Mangasarian–Fromovitz constraint qualification w.r.t. \(\partial\) is said to hold at \(\tilde{z} \in \mathcal{Z}\) whenever the condition

\[
0 \in \sum_{i \in I(\tilde{z})} \lambda_i \partial \varphi_i(\tilde{z}) + \sum_{i \in I} \lambda_i \left( \partial \varphi_i(\tilde{z}) \cup \partial \varphi_i(-\varphi_i)(\tilde{z}) \right), \quad \lambda_i \geq 0 \quad (i \in I(\tilde{z}) \cup J)
\]

is valid. Clearly, this reduces to the classical MFCQ when continuously differentiable functions \(\varphi_1, \ldots, \varphi_{p+q}\) are under consideration. Exploiting local boundedness and outer semicontinuity of the Clarke and limiting subdifferential, standard arguments show that \(\partial\)-NMFCQ is sufficient for \(A\partial\)-regularity. However, the study from [3, Section 4] for smooth functions clearly underlines that \(A\partial\)-regularity should be much weaker than \(\partial\)-NMFCQ in general. Below, we visualize this with the aid of two examples.

First, we want to review the nonsmooth variant of the relaxed constant positive linear dependence constraint qualification introduced in [24, Definition 1.1] via limiting normals.

**Definition 3.11:** Let \(\tilde{z} \in \mathcal{Z}\) be a feasible point of \((P)\) and assume that the functions \(\varphi_{p+1}, \ldots, \varphi_{p+q}\), which correspond to the equality constraints in \((P)\), are continuously differentiable in a neighbourhood of \(\tilde{z}\). We say that \(\partial\)-RCPLD, the relaxed constant positive linear dependence constraint qualification w.r.t. \(\partial\), holds at \(\tilde{z}\) whenever the following conditions are valid.

(i) The family \((\nabla \varphi_i(\tilde{z}))_{i \in I}\) has a constant rank in some neighbourhood of \(\tilde{z}\).

(ii) There is some index set \(\tilde{J} \subset I\) such that \((\nabla \varphi_i(\tilde{z}))_{i \in \tilde{J}}\) is a basis of the subspace \(\text{span}\{\nabla \varphi_i(\tilde{z}) \mid i \in \tilde{J}\}\).

(iii) For each index set \(\tilde{I} \subset I(\tilde{z})\) and each family of subgradients \((\xi_i)_{i \in \tilde{I}}\) satisfying \(\xi_i \in \partial \varphi_i(\tilde{z}) \quad (i \in \tilde{I})\) such that the pair of families \(((\xi_i)_{i \in \tilde{I}}, (\nabla \varphi_i(\tilde{z}))_{i \in \tilde{J}})\) is positive linearly dependent, we can ensure that, for large enough \(k \in \mathbb{N}\), the vectors from the family \((\xi_i^k)_{i \in \tilde{I}}\cup(\nabla \varphi_i(z^k))_{i \in \tilde{J}}\) are linearly dependent where the sequences \(\{z^k\}_{k \in \mathbb{N}} \subset \mathbb{R}^n\) and \(\{\xi_i^k\}_{k \in \mathbb{N}} \subset \mathbb{R}^n \quad (i \in \tilde{I})\) with \(z^k \to \tilde{z}\), \(\xi_i^k \to \xi_i \quad (i \in \tilde{I})\), and \(\xi_i^k \in \partial \varphi_i(z^k)\) for all \(k \in \mathbb{N}\) and \(i \in \tilde{I}\) are arbitrarily chosen.

By definition, we see that \(\partial\)-RCPLD is milder than \(\partial\)-NMFCQ whenever the equality constraints under consideration are smooth. Furthermore, \(\partial\)-RCPLD is sufficient for \(\partial\)-RCPLD. The converse is not true which can be exemplarily seen when considering the constraint region modelled by the single inequality constraint \(-|\cdot| \leq 0\) at \(\tilde{z} := 0\). Due to \(\partial(- |\cdot|)(0) =\)
\{−1, 1\} and \(\partial c(−|·|)(0) = \{−1, 1\}\), \(\partial\)-NMFCQ is valid which guarantees validity of \(\partial\)-RCPLD. On the other hand, \(\partial\)-RCPLD obviously fails to hold since 0 \(\notin \partial c(−|·|)(0)\) while \(\{1/k\}_{k \in \mathbb{N}} \subset \partial c(−|·|)(0)\) is non-vanishing and satisfies \(1/k \to 0\).

Let us mention that RCPLD has been introduced for standard nonlinear programs in [23], and assuming that all the functions \(\varphi_i\) \((i \in I \cup J)\) are smooth, Definition 3.11 recovers this classical notion. Recently, the definition of RCPLD has been extended to nonsmooth and complementarity-based systems in [24,34,35]. In [36], a parametric version of this constraint qualification has been discussed.

Below, we adapt the proofs of [[6, Theorem 4.8];[7, Theorem 4.2]] in order to verify that \(\partial □\)-RCPLD is indeed sufficient for \(A\partial □\)-regularity.

**Lemma 3.12:** Let \(\bar{z} \in \mathcal{Z}\) be a feasible point of \((P)\) and assume that the functions \(\varphi_{p+1}, \ldots, \varphi_{p+q}\) are continuously differentiable in a neighbourhood of \(\bar{z}\). Furthermore, let \(\partial □\)-RCPLD hold at \(\bar{z}\). Then \(\bar{z}\) is \(A\partial □\)-regular.

**Proof:** Let us fix some arbitrary point \(\eta \in \limsup_{z \to \bar{z}} \partial □(z)\). Then we find sequences \(\{z^k\}_{k \in \mathbb{N}}, \{\eta^k\}_{k \in \mathbb{N}} \subset \mathbb{R}^n\) such that \(z^k \to \bar{z}, \eta^k \to \eta, \) and \(\eta^k \in \partial □(z^k)\) for all \(k \in \mathbb{N}\). By construction, there are sequences \(\{\lambda^k_i\}_{k \in \mathbb{N}} \subset \mathbb{R}_+\) \((i \in I(\bar{z}))\), \(\{\xi^k_i\}_{k \in \mathbb{N}} \subset \mathbb{R}^n\) \((i \in I(\bar{z}))\), and \(\{\mu^k_i\}_{k \in \mathbb{N}} \subset \mathbb{R}^n\) \((i \in J)\) which satisfy \(\xi^k_i \in \partial □\varphi_i(z^k)\) for all \(i \in I(\bar{z})\) and \(k \in \mathbb{N}\) as well as

\[
\eta^k = \sum_{i \in I(\bar{z})} \lambda^k_i \xi^k_i + \sum_{i \in J} \mu^k_i \nabla \varphi_i(z^k)
\]

for all \(k \in \mathbb{N}\). Exploiting validity of \(\partial □\)-RCPLD (we use the notation from Definition 3.11), we find sequences \(\{\hat{\mu}^k_i\}_{k \in \mathbb{N}} \subset \mathbb{R}\) \((i \in \bar{J})\) such that

\[
\eta^k = \sum_{i \in I(\bar{z})} \lambda^k_i \xi^k_i + \sum_{i \in \bar{J}} \hat{\mu}^k_i \nabla \varphi_i(z^k)
\]

is valid for all \(k \in \mathbb{N}\). Next, for each \(k \in \mathbb{N}\), we apply [23, Lemma 1] in order to find an index set \(I^k \subset I(\bar{z})\) as well as multipliers \(\hat{\lambda}^k_i > 0\) \((i \in I^k)\) and \(\hat{\mu}^k_i \in \mathbb{R}\) \((i \in \bar{J})\) such that

\[
\eta^k = \sum_{i \in I^k} \hat{\lambda}^k_i \xi^k_i + \sum_{i \in \bar{J}} \hat{\mu}^k_i \nabla \varphi_i(z^k) \quad (8)
\]

while the vectors from \((\xi^k_i)_{i \in I^k} \cup (\nabla \varphi_i(z^k))_{i \in \bar{J}}\) are linearly independent. Noting that there are only finitely many subsets of \(I(\bar{z})\), we may assume w.l.o.g. that
Let us first show that $A\partial^\square$-regularity.

**Proof:** Let us first show that $\bar{z}$ is $A\partial^\square$-regular. Fix some $i \in I(\bar{z}) \cup J$. Since $\varphi_i$ is piecewise affine in a neighbourhood of $\bar{z}$, there only exist finitely many different regular (and, thus, limiting) normal cones to $\text{epi} \varphi_i$ locally around $\bar{z}$. Thus, we find a neighbourhood $U$ of $\bar{z}$ and finitely many closed sets $C^1, \ldots, C^\ell \subset \mathbb{R}^n$ such that, for each $z \in U$, there exists $v \in \{1, \ldots, \ell\}$ satisfying $\partial \varphi_i(z) = C^v$. Let us show $\partial \varphi_i(z) \subset \partial \varphi_i(\bar{z})$ for all $z \in V$ where $V \subset U$ is some neighbourhood of $\bar{z}$. Supposing that this is not true, we find a sequence $\{z^k\}_{k \in \mathbb{N}} \subset \mathbb{R}^n$ and some...
$v \in \{1, \ldots, \ell\}$ such that $z^k \to \bar{z}, \partial \varphi_i(z^k) = C^v$ for all $k \in \mathbb{N}$, and $C^v \setminus \partial \varphi_i(\bar{z}) \neq \emptyset$ hold. On the other hand, by outer semicontinuity of the limiting subdifferential and $z^k \to \bar{z}$, we automatically have $C^v \subset \partial \varphi_i(\bar{z})$, which is a contradiction.

Observing that there are only finitely many indices in $I(\bar{z}) \cup J$, we, thus, find a neighbourhood $\mathcal{U}$ of $\bar{z}$ such that $\partial \varphi_i(z) \subset \partial \varphi_i(\bar{z})$ holds for all $z \in \mathcal{U}$ and all $i \in I(\bar{z}) \cup J$ while $\partial (-\varphi_i)(z) \subset \partial (-\varphi_i)(\bar{z})$ is valid for all $z \in \mathcal{U}$ and all $i \in J$. Thus, we automatically have $\overline{\mathcal{M}}(z) \subset \overline{\mathcal{M}}(\bar{z})$ for all $z \in \mathcal{U}$, and this yields $A \partial$-regularity of $\bar{z}$.

Recalling that Clarke’s subdifferential corresponds to the closed convex hull of the limiting subdifferential, we infer $\mathcal{M}^{c}(z) \subset \mathcal{M}^{c}(\bar{z})$ for all $z \in \mathcal{U}$ from above, and this yields $A \partial^{c}$-regularity of $\bar{z}$.

Observe that in contrast to $\partial \Box$-RCPLD, we do not need to assume any smoothness of the functions $\varphi_{p+1}, \ldots, \varphi_{p+q}$, which model the equality constraints in (P), in Lemma 3.13. Furthermore, we would like to note that even in the situation where $\varphi_1, \ldots, \varphi_p$ are piecewise affine while $\varphi_{p+1}, \ldots, \varphi_{p+q}$ are affine, $\partial \Box$-RCPLD does not necessarily hold. In order to see this, one could simply consider the feasible set $\mathcal{Z}$ modelled by the single inequality constraint $|z| \leq 0$ at $\bar{z} := 0$. This is essentially different from the observations in standard nonlinear optimization where fully affine systems satisfy the constant rank constraint qualification and, thus, RCPLD (see [23, Section 3]).

The above observations underline that $A \partial \Box$-regularity is a very weak constraint qualification which may hold even for highly degenerated programs where standard constraint qualifications like $\partial \Box$-NMFCQ are always violated.

In Figure 1, we summarize the relations between the discussed constraint qualifications.

---

**Figure 1.** Relations between constraint qualifications addressing (P). The label $(\ast)$ indicates that the underlying relation is only reasonable whenever the functions $\varphi_{p+1}, \ldots, \varphi_{p+q}$ are continuously differentiable at the point of interest.
4. Application to mathematical programs with complementarity constraints

Here, we apply our results from Section 3 to complementarity constrained optimization problems and compare our findings with the ones from [6,7]. We focus on stationarity conditions and constraint qualifications.

For continuously differentiable data functions \( f: \mathbb{R}^n \to \mathbb{R}, \ g_i: \mathbb{R}^n \to \mathbb{R} \) \((i \in I^s := \{1, \ldots, m^s\})\), \( h_i: \mathbb{R}^n \to \mathbb{R} \) \((i \in I^h := \{1, \ldots, m^h\})\), and \( G_i, H_i: \mathbb{R}^n \to \mathbb{R} \) \((i \in I^{cc} := \{1, \ldots, m^{cc}\})\), where \( m^s, m^h \in \mathbb{N} \cup \{0\} \) and \( m^{cc} \in \mathbb{N} \) are arbitrary natural numbers, we consider the so-called mathematical program with complementarity constraints given by

\[
\min \left\{ f(z) \left| \begin{array}{l}
g_i(z) \leq 0 \ (i \in I^s), \ h_i(z) = 0 \ (i \in I^h), \\
0 \leq G_i(z) \perp H_i(z) \geq 0 \ (i \in I^{cc})
\end{array} \right. \right\}. \tag{MPCC}
\]

The latter has been studied quite intensively from the viewpoint of theory and numerical practice since this model covers numerous interesting real-world applications while being inherently irregular due to the challenging combinatorial structure of the feasible set (see, e.g. [16,17,38,39] and the references therein).

Throughout the section, the feasible set of (MPCC) will be denoted by \( \mathbb{Z}^{cc} \). For brevity of notation, let \( g: \mathbb{R}^n \to \mathbb{R}^{m^s} \), \( h: \mathbb{R}^n \to \mathbb{R}^{m^h} \), and \( G, H: \mathbb{R}^n \to \mathbb{R}^{m^{cc}} \) be the mappings which possess the component functions \( g_i \ (i \in I^s) \), \( h_i \ (i \in I^h) \), \( G_i \ (i \in I^{cc}) \), and \( H_i \ (i \in I^{cc}) \), respectively. Furthermore, for multipliers \( \lambda^s \in \mathbb{R}^{m^s} \), \( \lambda^h \in \mathbb{R}^{m^h} \), and \( \lambda^G, \lambda^H \in \mathbb{R}^{m^{cc}} \), we exploit the MPCC-Lagrangians

\[
\forall z \in \mathbb{R}^n: \quad L^{cc}_0(z, \lambda^s, \lambda^h, \lambda^G, \lambda^H) := (\lambda^s)^\top g(z) + (\lambda^h)^\top h(z) \\
+ (\lambda^G)^\top G(z) + (\lambda^H)^\top H(z)
\]

\[
L^{cc}(z, \lambda^s, \lambda^h, \lambda^G, \lambda^H) := f(z) + L^{cc}_0(z, \lambda^s, \lambda^h, \lambda^G, \lambda^H).
\]

For an arbitrary feasible point \( \bar{z} \in \mathbb{Z}^{cc} \), we define the following well-known index sets:

\[
I^{+0}(\bar{z}) := \{ i \in I^{cc} \mid G_i(\bar{z}) > 0 \land H_i(\bar{z}) = 0 \},
\]

\[
I^{0+}(\bar{z}) := \{ i \in I^{cc} \mid G_i(\bar{z}) = 0 \land H_i(\bar{z}) > 0 \},
\]

\[
I^{00}(\bar{z}) := \{ i \in I^{cc} \mid G_i(\bar{z}) = H_i(\bar{z}) = 0 \}.
\]

In order to apply the theory from Section 3 to the model (MPCC), we reformulate the final \( m^{cc} \) so-called complementarity constraints with the aid of the minimum function \( \varphi_{\min}: \mathbb{R}^2 \to \mathbb{R} \) given by \( \varphi_{\min}(a, b) := \min(a, b) \) for all \((a, b) \in \mathbb{R}^2\) in the following way:

\[
\varphi_{\min}(G_i(z), H_i(z)) = 0 \quad (i \in I^{cc}). \tag{9}
\]

Indeed, the resulting problem is equivalent to (MPCC) since \( \varphi_{\min} \) is a so-called nonlinear complementarity problem function (NCP-function for short) (see
Proposition 4.1: Let \( \bar{z} \in Z^{cc} \) be a feasible point of (MPCC). Then the following assertions hold.

(a) The point \( \bar{z} \) is \( A\partial c \)-stationary for (MPCC) using reformulation (9) if and only if there are sequences \( \{z^k\}_{k \in \mathbb{N}}, \{\epsilon^k\}_{k \in \mathbb{N}} \subset \mathbb{R}^n, \{\lambda^g,k\}_{k \in \mathbb{N}} \subset \mathbb{R}^{m^g}, \{\lambda^h,k\}_{k \in \mathbb{N}} \subset \mathbb{R}^{m^h} \) such that \( z^k \to \bar{z}, \epsilon^k \to 0, \) and

\[
\forall k \in \mathbb{N}: \quad \epsilon^k = \nabla_z L^{cc}(z^k, \lambda^g,k, \lambda^h,k, \lambda^{G,k}, \lambda^{H,k}), \quad (10a)
\]

\[
\forall k \in \mathbb{N}, \forall i \in I^g: \quad \min(\lambda_i^g,k, -g_i(\bar{z})) = 0, \quad (10b)
\]

\[
\forall k \in \mathbb{N}, \forall i \in I^{+0}(\bar{z}): \quad \lambda_i^{G,k} = 0, \quad (10c)
\]

\[
\forall k \in \mathbb{N}, \forall i \in I^{0+}(\bar{z}): \quad \lambda_i^{H,k} = 0, \quad (10d)
\]

\[
G_i(z^k) < H_i(z^k) \implies \lambda_i^{G,k} = 0, \quad (10e)
\]

Particularly, we find

\[
\forall k \in \mathbb{N}, \forall i \in I^{00}(\bar{z}): \quad (\lambda_i^{G,k})(\lambda_i^{H,k}) \geq 0 \quad (11)
\]

in this situation.

(b) The point \( \bar{z} \) is \( A\partial \)-stationary for (MPCC) using reformulation (9) if and only if there are sequences \( \{z^k\}_{k \in \mathbb{N}}, \{\epsilon^k\}_{k \in \mathbb{N}} \subset \mathbb{R}^n, \{\lambda^g,k\}_{k \in \mathbb{N}} \subset \mathbb{R}^{m^g}, \{\lambda^h,k\}_{k \in \mathbb{N}} \subset \mathbb{R}^{m^h} \) such that \( z^k \to \bar{z}, \epsilon^k \to 0, \) (10a)–(10d), and

\[
G_i(z^k) < H_i(z^k) \implies \lambda_i^{H,k} = 0, \quad (10f)
\]

\[
G_i(z^k) > H_i(z^k) \implies \lambda_i^{G,k} = 0, \quad (10g)
\]

\[
G_i(z^k) = H_i(z^k) \implies (\lambda_i^{G,k})(\lambda_i^{H,k}) \geq 0. \quad (10h)
\]

Particularly, we find

\[
\forall k \in \mathbb{N}, \forall i \in I^{00}(\bar{z}): \quad (\lambda_i^{G,k})(\lambda_i^{H,k}) \geq 0 \quad (12)
\]

in this situation. Above, \( \lor \) denotes the logical ‘or’. 

[40,41] for an overview). In order to evaluate the asymptotic stationarity and regularity conditions from Section 3 for (MPCC), based on reformulation (9) of the complementarity constraints, we exploit Lemma 2.1.
Proof: This follows by applying Definition 3.2 to the reformulated problem (MPCC) while respecting Lemma 2.1. On the route, one has to observe that for each sequence \( \{z^k\}_{k \in \mathbb{N}} \subset \mathbb{R}^n \) which satisfies \( z^k \to \bar{z} \), continuity of \( G \) and \( H \) yields the inclusions \( I^{0+}(\bar{z}) \subset \{ i \in I^{cc} \mid G_i(z^k) < H_i(z^k) \} \) as well as \( I^{+0}(\bar{z}) \subset \{ i \in I^{cc} \mid G_i(z^k) > H_i(z^k) \} \) for large enough \( k \in \mathbb{N} \).

Let us point out that our choice to use the NCP-function \( \varphi_{\min} \) for the reformulation of the complementarity constraints is essential in order to guarantee the validity of Proposition 4.1. Exemplary, let us mention that strictly weaker asymptotic stationarity conditions would be obtained if one exploits the famous Fischer–Burmeister function or the Kanzow–Schwartz function for that purpose. One nearby reason behind this fact is that these NCP-functions are actually too smooth. A related observation has been made in [22, Section 3] where or-constraints have been reformulated with the aid of so-called or-compatible NCP-functions.

In the light of Proposition 4.1, the following definition is reasonable.

Definition 4.2: Let \( \bar{z} \in \mathcal{Z}^{cc} \) be a feasible point of (MPCC).

(a) We call \( \bar{z} \) asymptotically MPCC-\( d^c \)-stationary (MPCC-A\( d^c \)-stationary) if there are sequences \( \{z^k\}_{k \in \mathbb{N}}, \{\varepsilon^k\}_{k \in \mathbb{N}} \subset \mathbb{R}^n, \{\lambda^g^k\}_{k \in \mathbb{N}} \subset \mathbb{R}^m, \{\lambda^h^k\}_{k \in \mathbb{N}} \subset \mathbb{R}^m \) as well as \( \{\lambda^G^k\}_{k \in \mathbb{N}}, \{\lambda^H^k\}_{k \in \mathbb{N}} \subset \mathbb{R}^{m^c} \) such that \( z^k \to \bar{z}, \varepsilon^k \to 0, (10a)–(10d), \) and (11) hold.

(b) We call \( \bar{z} \) asymptotically MPCC-\( p^c \)-stationary (MPCC-A\( p^c \)-stationary) if there are sequences \( \{z^k\}_{k \in \mathbb{N}}, \{\varepsilon^k\}_{k \in \mathbb{N}} \subset \mathbb{R}^n, \{\lambda^g^k\}_{k \in \mathbb{N}} \subset \mathbb{R}^m, \{\lambda^h^k\}_{k \in \mathbb{N}} \subset \mathbb{R}^m \) as well as \( \{\lambda^G^k\}_{k \in \mathbb{N}}, \{\lambda^H^k\}_{k \in \mathbb{N}} \subset \mathbb{R}^{m^c} \) such that \( z^k \to \bar{z}, \varepsilon^k \to 0, (10a)–(10d), \) and (12) hold.

Observe that these stationarity conditions generalize the classical notions of C- and M-stationarity of (MPCC) which are used throughout the literature (see, e.g. [39, Section 2.2] for the precise definitions).

Let us compare the conditions from Definition 4.2 with the sequential stationarity notions for (MPCC) which were studied in [6,7], respectively. First, we focus on the stationarity conditions w.r.t. the limiting subdifferential. Therefore, we introduce a set \( \Lambda^{cc} := \{(a,b) \in \mathbb{R}^2 \mid 0 \leq a \perp b \geq 0 \} \) and observe that conditions (10c), (10d), and (12) are equivalent to

\[
\forall k \in \mathbb{N} \forall i \in I^{cc}: (\lambda^G_i, \lambda^H_i) \in \mathcal{N}_{\Lambda^{cc}}(G_i(\bar{z}), H_i(\bar{z})).
\]

Thus, MPCC-A\( d^c \)-stationary points are MPCC-AKKT points in the sense of [7, Definition 3.2]. Exploiting the inclusion \( \mathcal{N}_{\Lambda^{cc}}(a,b) \subset \mathcal{N}_{\Lambda^{cc}}(\bar{a}, \bar{b}) \), which, due to the disjunctive structure of \( \Lambda^{cc} \), holds for all \((a,b) \in \Lambda^{cc}\) and all \((a,b) \in \Lambda^{cc} \cap U\) where \( U \subset \mathbb{R}^2 \) is a sufficiently small neighbourhood of \((\bar{a}, \bar{b})\), the converse relation holds true as well. Thus, the comments at the end of [6, Section 3] justify

\[\blacksquare\]
that MPCC-$A\partial$-stationarity corresponds to AM-stationarity in the sense of [6, Definition 3.3]. In similar way, we see that MPCC-$A\partial$-$c$-stationarity corresponds to AC-stationarity from [6, Definition 3.3]. Our approach, thus, recovers the sequential stationarity concepts from [6,7].

Clearly, due to Theorem 3.4, each local minimizer of (MPCC) is MPCC-$A\partial$-stationary. Observing that MPCC-$A\partial$-$\square$-stationarity is slightly weaker than $A\partial$-$\square$-stationarity for (MPCC) based on reformulation (9) of the complementarity constraints, we need a slightly stronger constraint qualification than $A\partial$-$\square$-regularity in order to infer the validity of the classical C- or M-stationarity conditions. More precisely, for some fixed feasible point $\bar{z} \in Z^{cc}$, one would be tempted to postulate outer semicontinuity of the mappings $\overline{M}^{cc}_{cc}, \overline{M}_{cc}: \mathbb{R}^n \Rightarrow \mathbb{R}^n$ given below for each $z \in \mathbb{R}^n$:

$$
\overline{M}^{cc}_{cc}(z) := \begin{cases} 
\nabla_z L^c_0(z, \lambda^g, \lambda^h, \lambda^G, \lambda^H) & \forall i \in I^g: \min(\lambda^g_i, -g_i(\bar{z})) = 0, \\
\forall i \in I^{0^+}(\bar{z}): \lambda^G_i = 0, \\
\forall i \in I^{0+}(\bar{z}): \lambda^H_i = 0, \\
\forall i \in I^{0^0}(\bar{z}): (\lambda^G_i)(\lambda^H_i) \geq 0
\end{cases}
$$

$$
\overline{M}_{cc}(z) := \begin{cases} 
\nabla_z L^c_0(z, \lambda^g, \lambda^h, \lambda^G, \lambda^H) & \forall i \in I^g: \min(\lambda^g_i, -g_i(\bar{z})) = 0, \\
\forall i \in I^{0^+}(\bar{z}): \lambda^G_i = 0, \\
\forall i \in I^{0+}(\bar{z}): \lambda^H_i = 0, \\
\forall i \in I^{0^0}(\bar{z}): \lambda^G_i, \lambda^H_i < 0 \\
\vee (\lambda^G_i)(\lambda^H_i) = 0
\end{cases}
$$

Observe that this precisely recovers the notions of AC- and AM-regularity from [6, Definition 4.1]. Employing [13, Theorem 5.3], outer semicontinuity of $\overline{M}^{cc}_{cc}$ equals the concept of AM-regularity from [13, Definition 3.8] applied to (MPCC) where the complementarity constraints are reformulated as

$$(G_i(z), H_i(z)) \in \Lambda^{cc}_{cc} (i \in I^{cc}),$$

and this, finally, is the same as the condition MPCC-CCP introduced in [7, Definition 3.9]. Note that the condition

$$\limsup_{z \to \bar{z}} \overline{M}^{cc}_{cc}(z) \subset \overline{M}^{cc}_{cc}(\bar{z}),$$

which, for example, holds in Example 3.8, amounts to a new constraint qualification for (MPCC) which guarantees C-stationarity of local minimizers. It is motivated by the more general concept of $wA\partial$-regularity from Definition 3.7.

Some algorithmic benefits of the above MPCC-tailored notions of asymptotic stationarity and regularity have been envisioned in [6,7]. More precisely, let us mention that the convergence analysis associated with some MPCC-tailored
penalty, multiplier-penalty, and relaxation methods can be carried out in the presence of sequential regularity.

The subsequent remark, which closes this section, underlines that the above ideas can be adapted in order to tackle other problem classes from disjunctive optimization.

**Remark 4.3:** For continuously differentiable functions \( \tilde{G}_i, \tilde{H}_i : \mathbb{R}^n \rightarrow \mathbb{R} \) \( (i \in I^{dc} := \{1, \ldots, m^{dc}\}) \) where \( m^{dc} \in \mathbb{N} \) is an arbitrary natural number, the model

\[
\min \left\{ f(z) \left| \begin{array}{l}
  g_i(z) \leq 0 \ (i \in I^g), \\
  h_i(z) = 0 \ (i \in I^h), \\
  \tilde{G}_i(z) \leq 0 \lor \tilde{H}_i(z) \leq 0 \ (i \in I^{dc})
\end{array} \right. \right\}
\]  

(MPOC)

is referred to as a mathematical program with or-constraints in the literature (see [21,22]). It basically suffers from the same problems as (MPCC) which is why weak stationarity notions and constraint qualifications as well as problem-tailored solution methods have been studied for (MPOC). Reformulating the final \( m^{dc} \) so-called or-constraints with the aid of

\[ \varphi_{\min}(\tilde{G}_i(z), \tilde{H}_i(z)) \leq 0 \quad (i \in I^{dc}), \]

which has also been suggested in [22, Section 3], we can apply the theory of Section 3 similarly as above in order to derive asymptotic stationarity and regularity conditions for (MPOC). More precisely, the resulting concepts of \( A\partial^c \)- and \( A\partial \)-stationarity generalize the notions of weak and M-stationarity for (MPOC) which have been introduced in [21, Definition 7.1].

Replacing the final \( m^{dc} \) constraints of (MPOC) by

\[ \tilde{H}_i(z) \geq 0, \quad \tilde{G}_i(z) \tilde{H}_i(z) \leq 0 \quad (i \in I^{dc}), \]

(VC)

a so-called vanishing constrained optimization problem is obtained (see, e.g. [18–20] and the references therein for an overview). Introducing the globally Lipschitz continuous function \( \varphi_{vc} : \mathbb{R}^2 \rightarrow \mathbb{R} \) by means of

\[
\forall (a,b) \in \mathbb{R}^2 : \varphi_{vc}(a,b) := \begin{cases} 
  a & a \geq 0, \ b \geq a, \\
  0 & a < 0, \ b \geq 0, \\
  |b| & \text{otherwise},
\end{cases}
\]

one can easily check that constraint system (VC) is equivalent to

\[ \varphi_{vc}(\tilde{G}_i(z), \tilde{H}_i(z)) \leq 0 \quad (i \in I^{dc}). \]

Exploiting the rules of subdifferential calculus, one can check that \( \partial^c \)- and \( \partial \)-stationarity of the associated problem (P) recover the M- and weak stationarity system of the vanishing constrained optimization problem (see [20, Definition 2.4] for the precise definitions). Consequently, we can proceed as
above in order to derive reasonable notions of asymptotic stationarity and regularity for vanishing constrained optimization problems.

Deriving the details in both cases is left to the interested reader.

Finally, let us note that the results of this section can be extended to disjunctive programs with nonsmooth data functions (see, e.g. [42–44] for applications, stationarity conditions, and constraint qualifications associated with nonsmooth complementarity and vanishing constrained optimization problems). We would like to point the reader’s attention to the fact that reformulating nonsmooth complementarity constraints with the aid of \( \varphi_{\min} \) yields equality constraints of type (9) again, but the subdifferentials of the mappings \( z \mapsto \varphi_{\min}(G_i(z), H_i(z)) \) \( i \in I_{cc} \) are now much more difficult to evaluate in general since \( G_i \) and \( H_i \) are potentially nonsmooth. Most likely, one might only be in a position to compute upper estimates of these subdifferentials with the aid of suitable chain rules (Lemma 2.1 does not apply anymore), i.e. a result similar to Proposition 4.1 would yield slightly weaker conditions than actual \( A \partial^c \) and \( A \partial \)-stationarity. Similar issues are likely to pop up when considering other types of nonsmooth disjunctive constraints.

5. Application to bilevel optimization problems with affine constraints

In this section, we suggest a solution method for the numerical handling of the optimistic bilevel optimization problem

\[
\min_{x,y} \{ f(x, y) \mid Cx + Dy \leq d, y \in \Psi(x) \}, \quad \text{(BPP)}
\]

where \( f: \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \) is continuously differentiable and \( \Psi: \mathbb{R}^n \to \mathbb{R}^m \) is the solution mapping of the fully linear parametric optimization problem

\[
\min_{y} \{ c^T y \mid Ax + By \leq b \}, \quad \text{(P(x))}
\]

i.e. for each \( x \in \mathbb{R}^n \), \( \Psi \) assigns to \( x \) the (possibly empty) solution set \( \Psi(x) \) of \( \text{(P(x))} \). Above, \( A \in \mathbb{R}^{p \times n}, B \in \mathbb{R}^{p \times m}, C \in \mathbb{R}^{q \times n}, D \in \mathbb{R}^{q \times m}, b \in \mathbb{R}^p, c \in \mathbb{R}^m, \) and \( d \in \mathbb{R}^q \) are fixed matrices. We refer to \( \text{(BPP)} \) and \( \text{(P(x))} \) as the upper and lower level problem, respectively. Although the constraints \( Cx + Dy \leq d \) as well as the lower level problem \( \text{(P(x))} \) are fully affine, \( \text{(BPP)} \) possesses a nonconvex feasible set which is only implicitly given. This makes \( \text{(BPP)} \) notoriously difficult even if \( f \) is fully linear. Besides, the model \( \text{(BPP)} \) covers numerous interesting applications such as inverse linear programming, balancing in energy and traffic networks, or data compression (see [25–27] and the references therein for an introduction to and a satisfying overview of bilevel optimization).
Subsequently, we want to exploit the so-called optimal value or marginal function \( \vartheta : \mathbb{R}^n \to \mathbb{R} \) of \((P(x))\) given by
\[
\forall x \in \mathbb{R}^n: \vartheta(x) := \inf_{y} \{ c^\top y \mid Ax + By \leq b \}.
\]

Due to full linearity of \((P(x))\), \( \vartheta \) is a convex and piecewise affine function. A fully explicit formula for its subdifferential can be found, e.g. in [45, Proposition 4.1].

**Lemma 5.1:** For each \( \bar{x} \in \text{dom} \vartheta \), the formula
\[
\partial \vartheta (\bar{x}) = A^\top S(\bar{x})
\]
holds, where \( S(\bar{x}) \) is the solution set of the dual problem associated with \((P(\bar{x}))\) given by
\[
S(\bar{x}) := \arg\max_\lambda \{(A\bar{x} - b)^\top \lambda \mid B^\top \lambda = -c, \lambda \geq 0\}.
\]

In this section, we exploit the well-known observation that \((BPP)\) is equivalent to
\[
\min_{x, y} \{ f(x, y) \mid Ax + By \leq b, \ c^\top y - \vartheta(x) \leq 0, \ Cx + Dy \leq d \} \quad \text{(VFR)}
\]
which is referred to as the value function reformulation of \((BPP)\). Observing that the estimate \( c^\top y - \vartheta(x) \geq 0 \) holds for each pair \((x, y) \in \mathbb{R}^n \times \mathbb{R}^m\) which satisfies \( Ax + By \leq b \), the constraint \( c^\top y - \vartheta(x) \leq 0 \) is actually active at each feasible point of \((VFR)\). One can check that this causes that, exemplarily, \( \vartheta^c \)-NMFCQ cannot hold at the feasible points of \((VFR)\) (see, e.g. [46, Proposition 3.2]). Besides, \( \vartheta \) is an implicitly given function whose full computation is not possible as soon as practically relevant problems are under consideration. Nevertheless, starting with [47], \((VFR)\) has been used successfully in the literature in order to derive optimality conditions and solution methods for \((BPP)\).

### 5.1. Stationarity conditions

Noting that the optimal value function \( \vartheta \) is locally Lipschitzian at each point from \( \text{int dom} \vartheta \), the following result is an immediate consequence of Lemma 3.13.

**Proposition 5.2:** Let \((\bar{x}, \bar{y}) \in \mathbb{R}^n \times \mathbb{R}^m\) be a local minimizer of \((BPP)\) and assume that \( \bar{x} \in \text{int dom} \vartheta \) holds. Then we find multipliers \( \lambda, \lambda' \in \mathbb{R}^p \) and \( \mu \in \mathbb{R}^q \) as well as \( \sigma \geq 0 \) such that
\[
\begin{align*}
0 &= \nabla_x f(\bar{x}, \bar{y}) + A^\top (\lambda - \sigma \lambda') + C^\top \mu, \quad (13a) \\
0 &= \nabla_y f(\bar{x}, \bar{y}) + B^\top (\lambda - \sigma \lambda') + D^\top \mu, \quad (13b) \\
0 &\leq \lambda \perp b - A\bar{x} - B\bar{y}, \quad (13c)
\end{align*}
\]
\[ 0 \leq \mu \perp d - Cx - Dy, \quad \text{(13d)} \]
\[ 0 = c + B^\top \lambda', \quad 0 \leq \lambda' \perp b - Ax - By. \quad \text{(13e)} \]

**Proof:** By assumption, \((\bar{x}, \bar{y})\) is a local minimizer of \((VFR)\) which is a Lipschitzian optimization problem in some neighbourhood of this point. Furthermore, the constraints of \((VFR)\) are given in terms of piecewise affine data functions. Applying Theorem 3.10 and Lemma 3.13, \((\bar{x}, \bar{y})\) is a \(\partial^c\)-stationary point of \((VFR)\). Thus, we find multipliers \(\lambda \in \mathbb{R}^p, \mu \in \mathbb{R}^q, \) and \(\sigma \geq 0\) as well as \(\xi \in \partial^c(-\vartheta)(\bar{x})\) satisfying

\[ 0 = \nabla_x f(\bar{x}, \bar{y}) + A^\top \lambda + \sigma \xi + C^\top \mu, \]
\[ 0 = \nabla_y f(\bar{x}, \bar{y}) + B^\top \lambda + \sigma c + D^\top \mu, \quad \text{(14)} \]
and (13c) as well as (13d). Observe that we have \(\partial^c(-\vartheta)(\bar{x}) = -\partial^c \vartheta(\bar{x}) = -\partial \vartheta(\bar{x})\) by convexity of \(\vartheta\). Due to Lemma 5.1 and strong duality of linear programming, we find \(\lambda' \in \mathbb{R}^p\) satisfying (13e) and \(\xi = -A^\top \lambda'.\) Inserting this and \(c = -B^\top \lambda'\) into (14) yields the claim. \(\blacksquare\)

Let us point out some important facts regarding the above result. First, under the assumptions made, \((\bar{x}, \bar{y})\) is already a \(\partial\)-stationary point of \((VFR)\). However, in order to evaluate \(\partial(-\vartheta)(\bar{x})\) successfully via Lemma 5.1, we have to estimate this set from above by \(\partial^c(-\vartheta)(\bar{x})\) and, again, end up with \(\partial^c\)-stationarity. This observation already has been made in [48, Sections 3 and 4] where a more general problem class has been considered.

Note that although the proof of Proposition 5.2 via Theorem 3.10 is novel, the result is well known in the literature on bilevel optimization. Indeed, the special structure of the lower level problem \((P(x))\) implies that \((VFR)\) is so-called partially calm at \((\bar{x}, \bar{y})\) (see [[46, Definition 3.1];[49, Theorem 4.1]] for details), which equivalently means that there is some \(\sigma \geq 0\) such that \((\bar{x}, \bar{y})\) is a local minimizer of

\[ \min_{x,y} \{ f(x, y) + \sigma(c^\top y - \vartheta(\bar{x})) \mid Ax + By \leq b, Cx + Dy \leq d \}, \]

and this problem can be tackled with standard KKT-theory since its constraints are affine. This approach precisely recovers stationarity system (13). Related approaches have been used, e.g. in [48,50,51] in order to derive necessary optimality conditions for more general bilevel optimization problems. Finally, we refer the interested reader to [52, Corollary 4.1] where yet another proof of Proposition 5.2 can be found which does not rely on the concept of partial calmness as well.

Note that the concept of asymptotic regularity is also applicable to more general bilevel optimization problems via their optimal value reformulation as long as the associated lower level optimal value function is locally Lipschitz continuous in a neighbourhood of the reference point. Keeping the weakness of
asymptotic regularity in mind, this approach could be suitable to find constraint qualifications which actually apply in a reasonable way to bilevel optimization problems. Further potentially promising constraint qualifications can be obtained when applying these concepts to the combined reformulation of the bilevel optimization problem where lower level value function and necessary optimality conditions are used in parallel (see [53]). A detailed investigation of these approaches is, however, beyond this paper’s scope.

5.2. A penalty-DC-method and its convergence properties

In the remainder of this section, we assume that $f = f_1 - f_2$ holds where the functions $f_1, f_2 : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ are continuously differentiable and convex, i.e. that $f$ is a so-called DC-function (see [28,29] for an overview of DC-optimization). This implies that (VFR) is a DC-problem, i.e. its objective function as well as its constraints are DC-functions. Below, we suggest a simple algorithm for the numerical solution of (BPP) which makes use of this observation while utilizing that the nonsmoothness is encapsulated only in $\vartheta$. Let us mention that these structural properties already have been used partially in the recent paper [54] for an algorithmic treatment of more general bilevel optimization problems where the data are allowed to be nonsmooth. More specifically and applied to our setting, the authors investigated the surrogate problem

$$
\min_{x,y} \{ f(x, y) \mid Ax + By \leq b, \ c^T y - \vartheta(x) \leq \varepsilon, \ Cx + Dy \leq d \}, \quad (VFR(\varepsilon))
$$

for some fixed relaxation parameter $\varepsilon \geq 0$, which is iteratively solved with some inexact DC-methods. In case $\varepsilon > 0$, a convergence theory is provided based on the observation that (VFR(\varepsilon)) behaves regularly in this case. However, the feasible set of (VFR(\varepsilon)) might be essentially larger than the one of (VFR) which is why this method most likely does not compute (asymptotically) feasible points of (BPP). Fixing $\varepsilon = 0$, no convergence guarantees were provided in [54]. We note that the authors of this paper did not investigate the situation where the relaxation parameter $\varepsilon$ is iteratively sent towards 0. We would like to mention [55] where DC-programming has been used to solve bilevel optimization problems of special structure to global optimality. However, in the latter paper, the authors do not exploit convexity of the optimal value function but make use of lower level optimality conditions to proceed. Yet another related approach for the global solution of bilevel optimization problems with fully convex lower level data can be found in [56,57].

Here, we stick a different path and investigate a penalty approach w.r.t. the constraint $c^T y - \vartheta(x) \leq 0$ where the resulting subproblems are solved with the aid of a DC-method. More precisely, exploiting the fact that the resulting penalized surrogate problems only possess affine constraints, they can be solved with the aid of the recently developed boosted DC-method from [58] in a reasonable
time. This method basically computes $\partial^c$-stationary points of DC-problems of type

$$\min_w \{ g(w) - h(w) \mid \tilde{A}w \leq \tilde{b} \}$$

where $g, h: \mathbb{R}^s \to \mathbb{R}$ are convex functions such that $g$ is continuously differentiable and $\tilde{A} \in \mathbb{R}^{t \times s}$ as well as $\tilde{b} \in \mathbb{R}^t$ are fixed matrices. For some fixed iterate $w^k \in \mathbb{R}^s$ and some subgradient $\xi^k \in \partial h(w^k)$, one identifies a minimizer $z^k \in \mathbb{R}^n$ of the partially linearized convex subproblem

$$\min \{ g(w) - (\xi^k)^\top w \mid \tilde{A}w \leq \tilde{b} \}$$

and performs a suitable line search along the direction $z^k - w^k$. For step size $\alpha^k$, the new iterate is set to $w^{k+1} := z^k + \alpha^k(z^k - w^k)$. The line search is responsible for the speed up. For details, we refer the interested reader to [58]. Note that setting $w^{k+1} := z^k$, i.e. choosing step size $\alpha^k := 0$, recovers the classical DC-method.

For brevity of notation, let us introduce $Z_\ell, Z_u \subset \mathbb{R}^n \times \mathbb{R}^m$ by means of

$$Z_\ell := \{ (x, y) \in \mathbb{R}^n \times \mathbb{R}^m \mid Ax + By \leq b \},$$
$$Z_u := \{ (x, y) \in \mathbb{R}^n \times \mathbb{R}^m \mid Cx + Dy \leq d \}.$$
**Algorithm 1** Penalty-DC-method for \((\text{BPP})\)

**Require:** \(\sigma^0 > 0, \gamma > 1, (x^0, y^0) \in Z_u \cap Z_\ell\)

1: Set \(k := 0.\)
2: Compute a \(\partial^c\)-stationary solution \((x^{k+1}, y^{k+1})\) of the DC-problem

\[
\min_{x, y} \begin{cases} f_1(x, y) + \sigma^k c^\top y - (f_2(x, y) + \sigma^k \vartheta(x)) \mid Ax + By \leq b; \\ Cx + Dy \leq d \end{cases}
\]

with the aid of the boosted DC-algorithm exploiting the starting point \((x^k, y^k)\).

3: if \(y^{k+1} \in \Psi(x^{k+1})\) then
4: Stop.
5: return \((x^{k+1}, y^{k+1})\)
6: else
7: Set \(\sigma^{k+1} := \gamma \sigma^k\) as well as \(k := k + 1\) and go to 2.
8: end if

**Proof:** Assume w.l.o.g. that \(x^k \to \bar{x}\) and \(y^k \to \bar{y}\) hold. By continuity of \(f\), the sequence \(f(x^{k+1}, y^{k+1})\) for \(k \in \mathbb{N}\) is bounded. Thus, we find \(\bar{k} > 0\) such that

\[
\forall k \in \mathbb{N}: 0 \leq c^\top y^{k+1} - \vartheta(x^{k+1}) \leq \bar{k}/\sigma^k.
\]

Taking the limit \(k \to \infty\) while exploiting the continuity of all appearing functions and \(\sigma^k \to \infty\), this shows \(c^\top \bar{y} - \vartheta(\bar{x}) = 0\), i.e. \(\bar{y} \in \Psi(\bar{x})\). This completes the proof.

Note that the boundedness assumption in Lemma 5.3 can be always guaranteed if the subproblems (16) are solved to global optimality. In this case, one can choose \(\kappa := f(\bar{x}, \bar{y})\) as an upper bound of the objective values where \((\bar{x}, \bar{y}) \in \mathbb{R}^n \times \mathbb{R}^m\) is an arbitrary feasible point of \((\text{BPP})\), and each accumulation point of \((x^k, y^k)\) is already a global minimizer of \((\text{BPP})\). However, one should note that solving the subproblems globally is, in general, only possible by decomposing the domain of \(\vartheta\) into its so-called regions of stability where \(\vartheta\) behaves in an affine way (see [56, Section 4] for a related approach) since this allows to trace back the solution of (16) to the solution of finitely many convex subproblems if \(f_2\) vanishes. Obviously, this approach is already computationally expensive whenever the dimension \(n\) is of medium size.

Now, we state a convergence result regarding Algorithm 1.
Theorem 5.4: Assume that Algorithm 1 generates a sequence \( \{(x^k, y^k)\}_{k \in \mathbb{N}} \subset \mathbb{R}^n \times \mathbb{R}^m \), and let \((\bar{x}, \bar{y}) \in \mathbb{R}^n \times \mathbb{R}^m \) be an accumulation point of this sequence which is feasible to (BPP). Then \((\bar{x}, \bar{y})\) is stationary for (BPP) in the sense of Proposition 5.2, i.e. there exist multipliers which solve stationarity system (13).

Proof: Let us assume w.l.o.g. that \(x^k \to \bar{x}\) and \(y^k \to \bar{y}\) hold. By construction of the method, for each \(k \in \mathbb{N}\), we find \(\xi^k \in \partial \varphi(x^{k+1})\), \(\lambda^k \in \mathbb{R}^p\), and \(\mu^k \in \mathbb{R}^q\) which satisfy

\[
0 = \nabla_x f(x^{k+1}, y^{k+1}) - \sigma^k \xi^k + A^\top \lambda^k + C^\top \mu^k, \quad (17a)
\]

\[
0 = \nabla_y f(x^{k+1}, y^{k+1}) + \sigma^k c + B^\top \lambda^k + D^\top \mu^k, \quad (17b)
\]

\[
0 \leq \lambda^k \perp b - Ax^{k+1} - By^{k+1}, \quad (17c)
\]

\[
0 \leq \mu^k \perp d - Cx^{k+1} - Dy^{k+1}. \quad (17d)
\]

Due to \(x^k \to \bar{x}\) and \(y^k \to \bar{y}\) as well as the feasibility of \((\bar{x}, \bar{y})\) for (BPP), we find that \((\bar{x}, \bar{y})\) is \(A\partial^c\)-stationary for (VFR). Recalling that each feasible point of (VFR) is \(A\partial^c\)-regular, \((\bar{x}, \bar{y})\) is already \(\partial^c\)-stationary for (VFR). As we have already mentioned before, (13) corresponds to the \(\partial^c\)-stationarity system of (VFR).

Let us note that the result of Theorem 5.4 remains true whenever the subproblems (16) are only solved up to approximate \(\partial^c\)-stationarity in the inner iteration, i.e. in terms of the more general DC-problem (15), we need to have

\[
\forall k \in \mathbb{N}: \; \epsilon^k \in \nabla g(w^{k+1}) - \partial h(w^{k+1}) + \tilde{A}^\top \xi^k, \quad 0 \leq \xi^k \perp \tilde{b} - \tilde{A} w^{k+1}
\]

such that \(\epsilon^k \to 0\) is guaranteed as \(k \to \infty\). Unluckily, the boosted DC-method from [58] does not guarantee this property of the outputs in general. Related phenomena motivated the study in [12] which, however, comes along with other drawbacks.

5.3. Implementation and numerical experiments

In this section, we are going to provide some numerical results regarding the computational competitiveness of Algorithm 1. Therefore, we compare the suggested method with two other intuitive penalty approaches which can be used to solve (BPP). On the one hand, it might be more natural to exploit the standard DC-algorithm (without boosting) to solve the subproblems (16) in Algorithm 1 which leads to a simpler method and speeds up the inner DC-iterations since no step size computation is necessary. On the other hand, it might be reasonable to
replace (16) by means of

\[
\min_{x,y,u} \left\{ f(x, y) + \sigma^k (c^\top y - (Ax - b)^\top u) \begin{array}{l}
Ax + By \leq b, \ Cx + Dy \leq d, \\
B^\top u = -c, \ u \geq 0
\end{array} \right\}
\]

(18)

in Algorithm 1 where the appearing penalty term models the duality gap of the lower level problem. This idea is basically taken from [59]. Observe that (18) is, in contrast to (16), a fully explicit optimization problem which can be solved via standard methods from constrained optimization. However, it does not possess the natural DC-structure we observed in (16). Furthermore, we are in need to treat the lower level Lagrange multiplier (or dual variable) \(u\) as an explicit variable which might be a delicate issue since (18) is related to a penalized version of the KKT-reformulation of the bilevel optimization problem (BPP). More precisely, the presence of \(u\) could induce artificial local minimizers and stationary points of (18) which do not correspond to local minimizers or stationary points of (16) (see [60,61] for details on this phenomenon).

Here, we challenge these three penalty methods with the aid of three bilevel optimization problems:

(a) the fully linear bilevel optimization problem from [62, Example 2],
(b) the problem from [63, Example 3.3] which possesses a quadratic upper level objective function, and
(c) an inverse transportation problem where the offer has to be reconstructed from a noisy transportation plan.

For each of these examples, we first provide a description of the problem data. Afterwards, we present our numerical results. More precisely, we challenge the three methods with random starting points and compare the outcome by means of computed function values, number of (outer) penalty iterations, number of (inner) DC-iterations (only for the DC-type methods), and the size of the (lower level) duality gap at the computed solution. In order to provide a reasonable visually convincing quantitative comparison, we make use of performance profiles (see [64]).

In the remainder of this section, we first comment on the actual numerical implementation of the algorithms. Afterwards, our results are presented.

5.3.1. **Implementation**

For a numerical comparison, we implemented the following penalty methods for the computational treatment of (BPP):

**PBDC**: Algorithm 1 where subproblems (16) are solved with the aid of the boosted DC-method from [58],
PDC: Algorithm 1 where the subproblems (16) are solved with the standard DC-method, and
PDG: in each iteration, we solve the subproblem (18) instead of (16) in Algorithm 1.

All these methods have been implemented using MATLAB 2021a. For the solution of the appearing subproblems, we employed MATLAB’s fmincon in default mode. Furthermore, appearing linear optimization problems, e.g. for the pointwise evaluation of the value function \( \vartheta \) or its subdifferential, see Lemma 5.1, are solved via MATLAB’s linprog routine. The sequence of penalty parameters is generated via \( \sigma_0 := 1 \) and \( \gamma := 1.2 \). Each of the algorithms is terminated whenever we have \( |c^\top(y^{k+1} - y^k)| \leq 10^{-7} \) for an arbitrary lower level solution \( y^{k+1} \in \Psi(x^{k+1}) \) or the number of outer iterations exceeds 200 (the latter, actually, did not happen). Note that this weakens the actual termination criterion \( y^{k+1} \in \Psi(x^{k+1}) \) which was used in Algorithm 1. For PBDC and PDC, we limited the number of inner DC-iterations to 100 (we never hit this bound during our experiments). Furthermore, using the notation from Section 5.2, we exploited the standard termination criterion \( \|z^k - w^k\| \leq 10^{-4} \) for both DC-methods. According to [58], the parameters for the line search in the boosted DC-method are fixed to \( \lambda := 1, \alpha := 10^{-2}, \) and \( \beta := 10^{-1} \). In order to enhance the numerical performance of the DC-methods PBDC and PDC, we added the zero \( \frac{1}{2} (\|x\|^2 + \|y\|^2) - \frac{1}{2} (\|x\|^2 + \|y\|^2) \) to the objective function of (16) in order to make both convex parts of it strongly convex.

All methods were challenged by 100 random starting points. These were generated by firstly choosing random vectors from a suitable box which are secondly projected onto the polyhedron \( \mathcal{Z}_u \cap \mathcal{Z}_\ell \). For the method PDG, we additionally chose \( u^0 \) as a solution of the linear optimization problem \( \min_u \{ e^\top u \mid B^\top u = -c, \ u \geq 0 \} \), where \( e \in \mathbb{R}^p \) denotes the all-ones-vector, and for each \( k \in \mathbb{N} \), used \( (x^k, y^k, u^k) \) as the initial point for the numerical solution of subproblem (18).

Finally, let us briefly comment on the creation of performance profiles. For the set \( \mathcal{A} := \{ \text{PBDC, PDC, PDG} \} \) of algorithms and the set \( \mathcal{S} := \{1, \ldots, \ell\} \) of indices associated with random starting points, let \( w^s_a \) be the output of algorithm \( a \in \mathcal{A} \) with random starting point \( s \in \mathcal{S} \). For a scalar performance measure \( \pi \) (representing computed function values, number of outer iterations, final lower level duality gap, or number of inner iterations), we consider the performance metric given by

\[
Q^\pi_{\theta}(w^s_a) := \begin{cases} 
\pi(w^s_a) - \pi^* + \theta & \text{if } w^s_a \text{ satisfies the termination criterion}, \\
\infty & \text{otherwise}.
\end{cases}
\]

Here, \( \pi^* \) represents a benchmark value which is chosen to be the globally optimal function value (or a suitable approximate of it) if \( \pi \) measures computed function values, and simply zero in the other three cases. Furthermore, \( \theta \geq 0 \) is an
additional parameter which reduces sensitivity to numerical accuracy and will be specified in the respective experiments. For the performance ratio

$$\forall a \in A \forall s \in S: r_{a,s}^\pi := \frac{Q^\pi_\theta (w^a_s)}{\min\{Q^\pi_\theta (w^a_{a'}) | a' \in A\}},$$

given for fixed $\pi$, we plot the illustrative parts of the curves $\rho_a^\pi : [1, \infty) \to [0, 1]$ ($a \in A$), defined by

$$\forall \tau \in [1, \infty): \rho_a^\pi (\tau) := \frac{\text{card}\{s \in S | r_{a,s}^\pi \leq \tau\}}{\text{card}(S)},$$

where $\text{card}(\cdot)$ assigns to each input set its cardinality.

### 5.3.2. Numerical experiments

**Experiment 1**: We investigate the linear bilevel optimization problem

$$\min_{x,y} \left\{ -2x_1 + x_2 + \frac{1}{2}y_1 \left| \begin{array}{c} x \geq 0, \ y \in \Psi(x) \end{array} \right. \right\}, \quad (\text{Ex1})$$

where $\Psi : \mathbb{R}^2 \Rightarrow \mathbb{R}^2$ is given by

$$\forall x \in \mathbb{R}^2 \colon \Psi(x) := \arg\min_y \left\{ -4y_1 + y_2 \left| \begin{array}{c} 2x_1 - y_1 + y_2 \geq \frac{5}{2}, \ x_1 + x_2 \leq 2, \\ x_1 - 3x_2 + y_2 \leq 2, \ y \geq 0 \end{array} \right. \right\},$$

which is taken from [62, Example 2]. For the creation of random starting points, we made use of the box $[0,2]^4$. The optimal objective value of this program is given by $f^* := -3.25$. The resulting performance profiles as well as some averaged numbers are presented in Table 1.

It turns out that **PBDC** as well as **PDG** reliably compute the global minimizer of (Ex1), and both methods do not outrun **PDC** in this regard. However, we see that **PDG** needs less outer iterations than the DC-methods until the termination criterion is reached. Nevertheless, we observe that **PBDC** needs less DC-iterations than **PDC**. Interestingly, the outputs of **PBDC** come along with a duality gap which is smaller by factor $10^{-1}$ than the upper bound $10^{-7}$ appearing in the termination criterion of the outer loop in several cases.

**Experiment 2**: Next, we investigate [63, Example 3.3] which is given by

$$\min_{x,y} \{x^2 + (y_1 + y_2)^2 | x \geq 0.5, \ y \in \Psi(x)\}, \quad (\text{Ex2})$$

where $\Psi : \mathbb{R} \Rightarrow \mathbb{R}^2$ is defined by

$$\forall x \in \mathbb{R} \colon \Psi(x) := \arg\min_y \{y_1 | x + y_1 + y_2 \geq 1, \ y \geq 0\}.$$ 

For the creation of random starting points, we exploited the box $[0,2]^3$. The optimal objective value of this program is given by $f^* := 0.5$. The resulting performance profiles as well as some averages regarding the performance indices can be found in Figure 3 and Table 2, respectively.
Figure 2. Performance profiles for (Ex1). From top left to bottom right: function values ($\theta = 10^{-4}$), outer iterations ($\theta = 1$), duality gap ($\theta = 10^{-6}$), and total inner iterations ($\theta = 1$).

Table 1. Averaged performance indices for (Ex1).

|                | PBDC   | PDC    | PDG    |
|----------------|--------|--------|--------|
| Average function value | $-3.2500$ | $-3.2214$ | $-3.2500$ |
| Average number of outer iterations | $12.4000$ | $11.2500$ | $7.8600$ |
| Average lower level duality gap | $8.9684 \cdot 10^{-9}$ | $4.0932 \cdot 10^{-8}$ | $4.7177 \cdot 10^{-8}$ |
| Average number of inner iterations | $5.1600$ | $9.6700$ | $-$ |

Similar to our first experiment, we observe that PBDC and PDG reliably compute the global minimizer. For that purpose, PBDC only needs to run one outer penalty iteration while PDG runs three outer iterations for each starting point. Regarding both criteria, PDC shows some flaws. Inspecting the total number of inner iterations, PBDC behaves much better than PDC. Finally, a look at the final duality gap shows that all three methods just fall below the upper bound $10^{-7}$ which was used in the termination criterion. Here, PDC seems to have slight advantages over the other two methods. However, it is already outperformed by the other two methods regarding the more important performance measures.

Experiment 3: Finally, we are going to challenge the three penalty methods by means of the inverse transportation problem

$$\min_{x,y} \left\{ \frac{1}{2} \|y - y_0\|^2 \right\} \quad \begin{array}{c} x \geq 0, \; e^\top x \geq e^\top b_{\text{dem}}, \; y \in \Psi(x) \end{array}, \quad (\text{Ex3})$$
Figure 3. Performance profiles for (Ex2). From top left to bottom right: function values ($\theta = 10^{-5}$), outer iterations ($\theta = 1$), duality gap ($\theta = 10^{-6}$), and total inner iterations ($\theta = 1$).

Table 2. Averaged performance indices for (Ex2).

|                | PBDC  | PDC   | PDG   |
|----------------|-------|-------|-------|
| Average function value | 0.5000 | 0.5285 | 0.5000 |
| Average number of outer iterations | 1.0000 | 4.8400 | 3.0000 |
| Average lower level duality gap | $7.0489 \cdot 10^{-8}$ | $1.5821 \cdot 10^{-8}$ | $5.5576 \cdot 10^{-8}$ |
| Average number of inner iterations | 4.7100 | 8.2700 | – |

where $\Psi : \mathbb{R}^n \Rightarrow \mathbb{R}^{n \times \ell}$ is the solution mapping of the parametric transportation problem

$$\min_y \left\{ \sum_{i=1}^n \sum_{j=1}^\ell c_{ij}y_{ij} \left| \begin{array}{l} \sum_{j=1}^\ell y_{ij} \leq x_i (i = 1, \ldots, n), \\
\sum_{i=1}^n y_{ij} \geq b_{\text{dem}}^j (j = 1, \ldots, \ell), \\
y \geq 0 \end{array} \right. \right\}. \quad (\text{TR}(x))$$

Above, $\ell \in \mathbb{N}$ is a positive integer, $b_{\text{dem}}^j \in \{0, \ldots, 10\}^\ell$ is a random integer vector which models the minimum demand of the $\ell$ consumers, and $c \in [0,1]^{n \times \ell}$ is a randomly chosen cost matrix. In (TR(x)), the parameter $x \in \mathbb{R}^n$ represents the offer provided at the $n$ warehouses which is unknown and shall be reconstructed.
Figure 4. Performance profiles for (Ex3). From top left to bottom right: function values ($\theta = 10^{-2}$), outer iterations ($\theta = 1$), duality gap ($\theta = 10^{-6}$), and total inner iterations ($\theta = 1$).

Table 3. Averaged performance indices for (Ex3).

|                | PBDC   | PDC    | PDG    |
|----------------|--------|--------|--------|
| Average function value | 0.5380 | 0.5561 | 0.2004 |
| Average number of outer iterations | 33.7200 | 22.3600 | 25.4200 |
| Average lower level duality gap | $2.4727 \cdot 10^{-8}$ | $3.2808 \cdot 10^{-8}$ | $6.2347 \cdot 10^{-8}$ |
| Average number of inner iterations | 177.2100 | 61.1100 | –       |

from a given (noised) transportation plan $y_0 \in \mathbb{R}^{n \times \ell}$. The latter is constructed in the following way: For $x_d := (e^\top b_{\text{dem}})/n \in \mathbb{R}$, we choose $y_d \in \Psi(x_d)$. Afterwards, some noise is added to $y_d$ in order to create $y_0$.

For our experiments, we chose $n := 5$ and $\ell := 7$. The precise values of the data $c$, $b_{\text{dem}}$, and $y_0$ used for our experiments can be found in the Appendix. The components of the random starting points are chosen from the interval $[0, 6]$. The minimum function value realized in our experiments is given by $f^* = 5.000776 \cdot 10^{-4}$. The associated point $(x^*, y^*)$ is also given in the Appendix. The resulting performance profiles and averaged performance indices are documented in Figure 4 and Table 3, respectively.

Clearly, (Ex3) is a far more challenging problem than (Ex1) or (Ex2). Throughout the test runs, we observed that the algorithms tend to identify different stationary points of (Ex3) with heavily differing objective values. However, in 87 of the 100 test runs, PBDC approximately identified $(x^*, y^*)$. In this regard, PDC
only succeeded in 10 of the test runs while PDG was successful in 71 of the test runs. Let us point out that the first row of Table 3 might be slightly misleading. Particularly, the surprisingly large value for the average function value related to PBDC results from the fact that the algorithm finds a point with objective value 2.9051 in 7 runs and a point with objective value 5.0909 in 5 runs. The performance advantage of PBDC regarding function values comes for a price, namely, a significantly larger number of outer and inner iterations in comparison with the other two methods. Regarding the size of the final duality gap, we did not figure out any surprising behaviour. Let us underline that although PDC comes along with the smallest number of outer and inner iterations, its outputs are often far away from \((x^*, y^*)\) which is why this advantage seems to be practically irrelevant.

5.3.3. Summary
Throughout the experiments, we observed that PBDC computes reasonable points, i.e. global minimizers, of (BPP) in most of the test runs. For the smaller test instances (Ex1) and (Ex2), some performance advantage regarding PDC w.r.t. inner and outer iteration numbers has been observed. In the more challenging setting of (Ex3), we abstain from putting too much emphasis on the iteration numbers of PDC since this algorithm did not compute points near the global minimizer in most of the runs. Furthermore, we attest PDG a solid performance regarding computed function values and iteration numbers. Here, the simplicity of subproblem (18) seems to pay off particularly for smaller problem instances.

6. Conclusions and perspectives
In this paper, we applied the concepts of asymptotic stationarity and regularity to nonlinear optimization problems with potentially nonsmooth but locally Lipschitzian data functions. Our theoretical investigations led to the formulation of three comparatively weak regularity conditions which enrich the landscape of available constraint qualifications in the field of nonsmooth programming (see Definition 3.7 and Figure 1). Afterwards, we investigated complementarity constrained programs in order to show that these quite general concepts possess some reasonable extensions to disjunctive programs where they can be used in order to carry out the convergence analysis associated with some solution methods under weak assumptions. Pointing the reader’s attention back to Remark 4.3, our results make clear that similar concepts can be easily obtained for or- and vanishing constrained optimization problems. It remains a task for future research to study the capability of asymptotic stationarity and regularity in the context of the numerical treatment of these problem classes. Finally, we exploited asymptotic regularity in the context of bilevel optimization. More precisely, we justified certain stationarity conditions as well as a penalty method for the numerical solution of affinely constrained bilevel optimization problems. Results of some computational experiments are shown in order to provide a quantitative justification
of our approach. We already pointed out that the overall theory can be easily extended to bilevel optimization problems with nonlinear but fully convex constraints. However, in this situation, asymptotic regularity is no longer inherently satisfied. It remains to be seen whether this concept yields applicable constraint qualifications for more general bilevel programming problems. Furthermore, it has to be studied whether the ideas behind Algorithm 1 still lead to convincing numerical results as soon as further nonlinearities appear in the problem data. Suitable test problems can be found in the BOLIB collection from [65] which also comprises (Ex1) and (Ex2) considered in Section 5.3.2.

**Disclosure statement**

No potential conflict of interest was reported by the author(s).

**ORCID**

Patrick Mehlitz [http://orcid.org/0000-0002-9355-850X](http://orcid.org/0000-0002-9355-850X)

**References**

[1] Andreani R, Martínez JM, Svaiter BF. A new sequential optimality condition for constrained optimization and algorithmic consequences. SIAM J Optim. 2010;20(6):3533–3554.

[2] Andreani R, Haeser G, Martínez JM. On sequential optimality conditions for smooth constrained optimization. Optimization. 2011;60(5):627–641.

[3] Andreani R, Martínez JM, Ramos A, et al. A cone-continuity constraint qualification and algorithmic consequences. SIAM J Optim. 2016;26(1):96–110.

[4] Andreani R, Martínez JM, Ramos A, et al. Strict constraint qualifications and sequential optimality conditions for constrained optimization. Math Oper Res. 2018;43(3):693–717.

[5] Andreani R, Fazzio NS, Schuverdt ML, et al. A sequential optimality condition related to the quasi-normality constraint qualification and its algorithmic consequences. SIAM J Optim. 2019;29(1):743–766.

[6] Andreani R, Haeser G, Secchin LD, et al. New sequential optimality conditions for mathematical programs with complementarity constraints and algorithmic consequences. SIAM J Optim. 2019;29(4):3201–3230.

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

[8] Kanzow C, Raharja AB, Schwartz A. Sequential optimality conditions for cardinality-constrained optimization problems with applications. Comput Optim Appl. 2021;80(1):185–211.

[9] Krulikovski EHM, Ribeiro AA, Sachine M. A sequential optimality condition for mathematical programs with cardinality constraints. Preprint arXiv, 2020, p. 1–23. Available from: https://arxiv.org/abs/2008.03158 (accessed on Jan 13, 2022).

[10] Andreani R, Haeser G, Viana DS. Optimality conditions and global convergence for nonlinear semidefinite programming. Math Program. 2020;180(1):203–235.
[11] Börgens E, Kanzow C, Mehlitz P, et al. New constraint qualifications for optimization problems in Banach spaces based on asymptotic KKT conditions. SIAM J Optim. 2020;30(4):2956–2982.
[12] Helou ES, Santos SA, Simões LEA. A new sequential optimality condition for constrained nonsmooth optimization. SIAM J Optim. 2020;30(2):1610–1637.
[13] Mehlitz P. Asymptotic stationarity and regularity for nonsmooth optimization problems. J Nonsmooth Anal Optim. 2020;1:6575.
[14] Clarke FH. Optimization and nonsmooth analysis. New York (NY): Wiley; 1983.
[15] Mordukhovich BS. Variational analysis and generalized differentiation, part I: basic theory, part II: applications. Berlin: Springer; 2006.
[16] Luo Z-Q, Pang J-S, Ralph D. Mathematical programs with equilibrium constraints. Cambridge: Cambridge University Press; 1996.
[17] Outrata JV, Kočvara M, Zowe J. Nonsmooth approach to optimization problems with equilibrium constraints. Dordrecht: Kluwer Academic; 1998.
[18] Achtziger W, Kanzow C. Mathematical programs with vanishing constraints: optimality conditions and constraint qualifications. Math Program. 2008;114(1):69–99.
[19] Hoheisel T, Kanzow C. First- and second-order optimality conditions for mathematical programs with vanishing constraints. Appl Math. 2007;52(6):495–514.
[20] Hoheisel T, Pablos B, Pooladian A, et al. A study of one-parameter regularization methods for mathematical programs with vanishing constraints. Optim Methods Softw. 2020;1–43.
[21] Mehlitz P. Stationarity conditions and constraint qualifications for mathematical programs with switching constraints. Math Program. 2020;181(1):149–186.
[22] Mehlitz P. A comparison of solution approaches for the numerical treatment of or-constrained optimization problems. Comput Optim Appl. 2020;76(1):233–275.
[23] Andreani R, Haeser G, Schuverdt ML, et al. A relaxed constant positive linear dependence constraint qualification and applications. Math Program. 2012;135(1):255–273.
[24] Xu M, Ye JJ. Relaxed constant positive linear dependence constraint qualification and its application to bilevel programs. J Global Optim. 2020;78:181–205.
[25] Dempe S. Foundations of bilevel programming. Dordrecht: Kluwer; 2002.
[26] Dempe S. Bilevel optimization: theory, algorithms, applications and a bibliography. In: Dempe S, Zemkoho AB, editors. Bilevel optimization: advances and next challenges. Cham: Springer International; 2020. p. 581–672.
[27] Dempe S, Kalashnikov V, Pérez-Valdéz G, et al. Bilevel programming problems – theory, algorithms and applications to energy networks. Berlin: Springer; 2015.
[28] An LTH, Tao PD. DC programming and DCA: thirty years of developments. Math Program. 2018;169(1):5–68.
[29] Horst R, Thoai NV. DC programming: overview. J Optim Theory Appl. 1999;103(1):1–43.
[30] Rockafellar RT, Wets RJ-B. Variational analysis. Berlin: Springer; 1998.
[31] Mehlitz P. On the linear independence constraint qualification in disjunctive programming. Optimization. 2020;69(10):2241–2277.
[32] Vinter R. Optimal control. New York (NY): Birkhäuser; 2000.
[33] Dutta J, Deb K, Tulshyan R, et al. Approximate KKT points and a proximity measure for termination. J Global Optim. 2013;56:1463–1499.
[34] Chieu NH, Lee GM. A relaxed constant positive linear dependence constraint qualification for mathematical programs with equilibrium constraints. J Optim Theory Appl. 2013;158(1):11–32.
[35] Guo L, Lin G-H. Notes on some constraint qualifications for mathematical programs with equilibrium constraints. J Optim Theory Appl. 2013;156:600–616.
[36] Mehlitz P, Minchenko LI. R-regularity of set-valued mappings under the relaxed constant positive linear dependence constraint qualification with applications to parametric and bilevel optimization. Set-valued Var Anal. 2021;1–27.

[37] Robinson SM. Some continuity properties of polyhedral multifunctions. In König H, Korte B, Ritter K, editors. Mathematical programming at Oberwolfach. Berlin: Springer; 1981. p. 206–214.

[38] Scheel S, Scholtes S. Mathematical programs with complementarity constraints: stationarity, optimality, and sensitivity. Math Oper Res. 2000;25(1):1–22.

[39] Ye JJ. Necessary and sufficient optimality conditions for mathematical programs with equilibrium constraints. J Math Anal Appl. 2005;307(1):350–369.

[40] Galántai A. Properties and construction of NCP functions. Comput Optim Appl. 2012;52(3):805–824.

[41] Kanzow C, Yamashita N, Fukushima M. New NCP-functions and their properties. J Optim Theory Appl. 1997;94(1):115–135.

[42] Kazemi S, Kanzi N. Constraint qualifications and stationarity conditions for mathematical programs with non-differentiable vanishing constraints. J Optim Theory Appl. 2018;179:800–819.

[43] Movahedian N, Nobakhtian S. Nondifferentiable multiplier rules for optimization problems with equilibrium constraints. J Convex Anal. 2009;16(1):187–210.

[44] Movahedian N, Nobakhtian S. Constraint qualifications for nonsmooth mathematical programs with equilibrium constraints. Set-Valued Var Anal. 2009;17:63–95.

[45] Ye JJ, Wu S-Y. First order optimality conditions for generalized semi-infinite programming problems. J Optim Theory Appl. 2008;137(2):419–434.

[46] Ye JJ, Zhu DL. Optimality conditions for bilevel programming problems. Optimization. 1995;33(1):9–27.

[47] Outrata JV. A note on the usage of nondifferentiable exact penalties in some special optimization problems. Kybernetika. 1988;24(4):251–258.

[48] Dempe S, Dutta J, Mordukhovich BS. New necessary optimality conditions in optimistic bilevel programming. Optimization. 2007;56(5–6):577–604.

[49] Mehlitz P, Minchenko LI, Zemkoho AB. A note on partial calmness for bilevel optimization problems with linearly structured lower level. Optim Lett. 2021;15:1277–1291.

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

[51] Mordukhovich BS, Nam NM, Phan HM. Variational analysis of marginal functions with applications to bilevel programming. J Optim Theory Appl. 2012;152(3):557–586.

[52] Ye JJ. Nondifferentiable multiplier rules for optimization and bilevel optimization problems. SIAM J Optim. 2004;15(1):252–274.

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

[54] Ye JJ, Yuan X, Zeng S, et al. Difference of convex algorithms for bilevel programs with applications in hyperparameter selection. Preprint arXiv, 2021. p. 1–29. Available from: https://arxiv.org/abs/2102.09006 (accessed on Jan 13, 2022).

[55] An LTH, Tao PD, Canh NN, et al. DC programming techniques for solving a class of nonlinear bilevel programs. J Global Optim. 2008;44(3):313–337.

[56] Dempe S, Franke S. On the solution of convex bilevel optimization problems. Comput Optim Appl. 2016;63(3):685–703.

[57] Dempe S, Harder F, Mehlitz P, et al. Solving inverse optimal control problems via value functions to global optimality. J Global Optim. 2019;74(2):297–325.
Aragón Artacho FJ, Campoy R, Vuong PT. The boosted DC algorithm for linearly constrained DC programming. Preprint arXiv, 2021. p. 1–24. Available from: https://arxiv.org/abs/1908.01138 (accessed on Jan 13, 2022).

White DJ, Anandalingam G. A penalty function approach for solving bi-level linear programs. J Global Optim. 1993;3:397–419.

Benko M, Mehlitz P. On implicit variables in optimization theory. J Nonsmooth Anal Optim. 2021;2:7215.

Dempe S, Dutta J. Is bilevel programming a special case of a mathematical program with complementarity constraints? Math Program. 2012;131(1):37–48.

Bard JF, Falk JE. An explicit solution to the multi-level programming problem. Comput Oper Res. 1982;9(1):77–100.

Lampariello L, Sagratella S. A bridge between bilevel programs and Nash games. J Optim Theory Appl. 2017;174:613–635.

Dolan ED, Moré JJ. Benchmarking optimization software with performance profiles. Math Program. 2002;91(2):201–213.

Zhou S, Zemkoho AB, Tin A. BOLIB: bilevel optimization library of test problems. In: Dempe S, Zemkoho AB, editors. Bilevel optimization: advances and next challenges. Cham: Springer International; 2020. p. 563–580.

Appendix

Here, we provide the missing data for Experiment 3 from Section 5.3.2. First, we state the data matrices $c \in [0, 1]^{5 \times 7}$, $b^{\text{dem}} \in \{1, \ldots, 10\}^7$, and $y_0 \in \mathbb{R}^{5 \times 7}$:

$$
c = \begin{pmatrix}
0.5757 & 0.8423 & 0.4997 & 0.4390 & 0.1491 & 0.0283 & 0.7567 \\
0.7961 & 0.2936 & 0.1152 & 0.3751 & 0.8289 & 0.8418 & 0.6652 \\
0.9601 & 0.9431 & 0.1127 & 0.6483 & 0.4808 & 0.0665 & 0.8978 \\
0.4972 & 0.7713 & 0.0604 & 0.2625 & 0.6511 & 0.0136 & 0.6385 \\
0.3849 & 0.7657 & 0.6529 & 0.3815 & 0.0300 & 0.3401 & 0.9189
\end{pmatrix},
$$

$$
b^{\text{dem}} = \begin{pmatrix}
5 & 5 & 5 & 10 & 3 & 9 & 1
\end{pmatrix}^T,
$$

$$
y_0 = \begin{pmatrix}
-0.0032 & 0.0053 & -0.0031 & 0.0024 & 2.9991 & 4.5902 & 0.0020 \\
0.0020 & 5.0030 & 1.5969 & -0.0001 & 0.0040 & 0.0078 & 0.9911 \\
-0.0080 & 0.0030 & 3.2053 & 0.0098 & -0.0075 & 4.3973 & 0.0035 \\
-0.0025 & 0.0073 & 0.1958 & 7.3927 & 0.0035 & -0.0059 & 0.0074 \\
5.0050 & -0.0016 & -0.0100 & 2.5930 & -0.0045 & 0.0074 & 0.0020
\end{pmatrix}.
$$

Next, we state the solution $(x^*, y^*) \in \mathbb{R}^5 \times \mathbb{R}^{5 \times 7}$ with the best function value $f^* = 5.000766 \cdot 10^{-4}$ found during our experiments:

$$
x^* = \begin{pmatrix}
7.5975 & 7.5975 & 7.6095 & 7.5964 & 7.6002
\end{pmatrix}^T,
$$

$$
y^* = \begin{pmatrix}
0 & 0 & 0 & 0 & 3.0000 & 4.5965 & 0 \\
0 & 5.0000 & 1.5975 & 0 & 0 & 0 & 1.0000 \\
0 & 0 & 3.2060 & 0 & 0 & 4.4035 & 0 \\
0 & 0 & 0.1965 & 7.3998 & 0 & 0 & 0 \\
5.0000 & 0 & 0 & 2.6002 & 0 & 0 & 0
\end{pmatrix}.
$$

We note that the desired pair of variables $(x_d, y_d) \in \mathbb{R}^5 \times \mathbb{R}^{5 \times 7}$, which has been used for the precise construction of the problem data, is given as stated below:

$$
x_d = \begin{pmatrix}
7.6000 & 7.6000 & 7.6000 & 7.6000 & 7.6000
\end{pmatrix}^T,
$$
\[ y_d = \begin{pmatrix}
0 & 0 & 0 & 0 & 3.0000 & 4.6000 & 0 \\
0 & 5.0000 & 1.6000 & 0 & 0 & 0 & 1.0000 \\
0 & 0 & 3.2000 & 0 & 0 & 4.4000 & 0 \\
0 & 0 & 0.2000 & 7.4000 & 0 & 0 & 0 \\
5.0000 & 0 & 0 & 2.6000 & 0 & 0 & 0 
\end{pmatrix}. \]