Yokohama Publishers
Online Journal for Editors

Pacific Journal of Optimization
Vol. 9, No. 2, pp. 183–199, April 2013

BILEVEL PROGRAMMING: STATIONARITY AND STABILITY

S. Dempe and S. Franke

Abstract: The bilevel programming problem and its Karush-Kuhn-Tucker reformulation as well as a nonlinear surrogate problem are considered. Appropriate constraint qualifications and stationarity concepts are investigated, the equivalence of two necessary optimality conditions is shown. Stability theorems are proved. Finally, properties of the optimal value function and the optimal set mapping of a nonlinear optimization problem related to the classical optimistic bilevel programming problem are investigated.

Key words: *bilevel programming, mathematical programs with equilibrium constraints, S-stationarity, M-stationarity, strong stability*

Mathematics Subject Classification: *90C30, 91A65, 90C31*

1 Introduction

Bilevel programming problems are hierarchical optimization problems between two participants, the follower and the leader. The leader first selects a vector x. This vector is a parameter in a second optimization problem which is solved by the follower. The follower selects a response $y(x)$ out of the set of optimal solutions and communicates this to the leader. Formally, this is described by

$$\begin{array}{ll} \min\limits_{x} & F(x,y) \\ \text{s.t.} & G(x) \leq 0 \\ & y \in \Psi(x) := \operatorname*{Argmin}\limits_{z} \{f(x,z) : \, g(x,z) \leq 0\}. \end{array}$$

Since $\Psi(x)$ can be a point-to-set-mapping, the response of the follower may not be clear. In order to avoid this ambiguity, we presuppose a cooperation between the two players which leads to the optimistic bilevel programming problem

$$\begin{array}{ll} \min\limits_{x,y} & F(x,y) \\ \text{s.t.} & G(x) \leq 0 \\ & y \in \Psi(x) := \operatorname*{Argmin}\limits_{z} \{f(x,z) : \, g(x,z) \leq 0\}. \end{array} \tag{1.1}$$

For differences between this and the pessimistic approach, the interested reader is referred to [1].

ISSN 1348-9151

We assume throughout the paper that $F, f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $G : \mathbb{R}^n \to \mathbb{R}^k$ and $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$ are continuously differentiable functions.

Additional equality constraints can easily be included in both upper and lower level of the bilevel programming problem. They can be handled by a slight modification of the constraint qualifications related to (1.1). But since the complementarity structures we want to examine stem from the inequality constraints, we dropped the equality constraints here for ease of reference.

There exist various ways to reformulate (1.1) into a single level problem. One possibility is the usage of the lower level parameter-dependent nondifferentiable optimal value function

$$\varphi(x) := \min_y \{f(x, y) : \, g(x, y) \le 0\}.$$

The resulting problem, which is fully equivalent to (1.1), has been extensively studied in [3, 24].

In this paper, we want to examine another approach. Therefore, let the objective function $f(x, \cdot)$ as well as the lower level constraints $g(x, \cdot) \le 0$ be convex for every x with $G(x) \le 0$. Furthermore, we assume that the lower level fulfills a certain constraint qualification like the Mangasarian-Fromovitz constraint qualification (MFCQ). Then it is possible to apply the Karush-Kuhn-Tucker (KKT) conditions to the lower level which leads to the following reformulation of the bilevel programming problem, the KKT reformulation

$$\begin{aligned} &\min_{x,y,\lambda} F(x, y) \\ &\text{s.t.} \begin{cases} G(x) \le 0, \; g(x, y) \le 0 \\ \lambda \ge 0, \; \lambda^\top g(x, y) = 0 \\ \mathcal{L}(x, y, \lambda) := \nabla_y f(x, y) + \nabla_y g(x, y)^\top \lambda = 0. \end{cases} \end{aligned} \tag{1.2}$$

Problem (1.2) is fully equivalent to the bilevel programming problem if global optimal solutions are considered [1]. However, this is not true for local optimal solutions, as was recently shown in [2]. Only if a point $(\overline{x}, \overline{y}, \lambda)$ is a local optimal solution of (1.2) for every feasible Lagrange multiplier

$$\lambda \in \Lambda(\overline{x}, \overline{y}) := \left\{ \lambda \in \mathbb{R}^p : \, \lambda_i \ge 0 \, (i = 1, ..., p), \, \mathcal{L}(\overline{x}, \overline{y}, \lambda) = 0, \, \lambda^\top g(\overline{x}, \overline{y}) = 0 \right\},$$

then $(\overline{x}, \overline{y})$ is also locally optimal for (1.1).

The programming problem (1.2) is referred to as classical KKT reformulation. The lower level reformulation via a generalized equation is called primal KKT approach; investigations regarding this problem can be found in [4].

Since (1.2) contains complementarity terms of the form $\lambda^\top g(x, y) = 0$, the problem relies on the theory of mathematical problems with equilibrium constraints (MPEC). A detailed analysis of this class of problems, including optimality conditions and solution algorithms, can be found in [14].

For a feasible point $(\overline{x}, \overline{y}, \overline{\lambda})$ of (1.2), we define the following index sets which will be the basis for our analysis:

$$\theta := \theta\left(\overline{x}, \overline{y}, \overline{\lambda}\right) = \{i : \ G_i(\overline{x}) = 0\}$$
$$\eta := \eta\left(\overline{x}, \overline{y}, \overline{\lambda}\right) = \{i : \ \overline{\lambda}_i = 0, \ g_i(\overline{x}, \overline{y}) < 0\}$$
$$\mu := \mu\left(\overline{x}, \overline{y}, \overline{\lambda}\right) = \{i : \ \overline{\lambda}_i = 0, \ g_i(\overline{x}, \overline{y}) = 0\}$$
$$\nu := \nu\left(\overline{x}, \overline{y}, \overline{\lambda}\right) = \{i : \ \overline{\lambda}_i > 0, \ g_i(\overline{x}, \overline{y}) = 0\}.$$

As we will see in Section 3, the classification of stationarity concepts strongly depends on the so-called degenerate set μ.

For a feasible point $(\overline{x}, \overline{y}, \overline{\lambda})$ of (1.2), let (A, B) be a partition of μ, i.e. $A \cap B = \emptyset$, $A \cup B = \mu$. Then, for every such pair (A, B), we can define the nonlinear programming problem

$$\min_{x,y,\lambda} F(x, y)$$
$$\text{s.t.} \begin{cases} G(x) \leq 0 \\ g_i(x, y) = 0, \ \lambda_i \geq 0 \quad i \in \nu \cup A \\ g_i(x, y) \leq 0, \ \lambda_i = 0 \quad i \in \eta \cup B \\ \mathcal{L}(x, y, \lambda) = 0. \end{cases} \tag{1.3}$$

One can easily deduce that the union over the feasible sets of the arising problems for all tuples (A, B) is equal to the feasible set of (1.2) in a sufficiently small neighborhood of the point $(\overline{x}, \overline{y}, \overline{\lambda})$. Unfortunately, the number of partitions equals $2^{|\mu|}$ and may become large if there exist many degenerate constraints.

The following notations are used throughout the paper. The closure of a set $M \subseteq \mathbb{R}^n$ is denoted by cl M. The graph of a point-to-set-mapping $\Psi : \mathbb{R}^n \to 2^{\mathbb{R}^m}$ reads

$$\text{gph}\ \Psi = \{(x, y) : \ y \in \Psi(x)\}.$$

For a differentiable function $F : \mathbb{R}^n \to \mathbb{R}$, $\nabla F(\overline{x}) = (\nabla_{x_1} F(\overline{x}), ..., \nabla_{x_n} F(\overline{x}))$ denotes the gradient of F at $\overline{x}$. $U(\overline{x}) := U_\varepsilon(\overline{x})$ denotes the open ball around $\overline{x}$ with a sufficiently small radius $\varepsilon > 0$. The unit ball is denoted by $\mathbb{B}$. The i-th component of a vector $a \in \mathbb{R}^n$, $a = (a_1, ..., a_n)^\top$ is denoted by a_i. For a matrix $A \in \mathbb{R}^{n \times m}$, the transposed matrix is $A^\top$.

2 Foundations and Definitions

In this section, some background material which is needed in the paper is provided.

Definition 2.1. Let $F : \mathbb{R}^n \to \mathbb{R}$ be lower semicontinuous at $\overline{x}$. The proximal subdifferential of F at $\overline{x}$ is defined by

$$\partial^\pi F(\overline{x}) = \{ \ d \in \mathbb{R}^n : \exists\, L > 0, \varepsilon > 0 \text{ with}$$
$$F(x) \geq F(\overline{x}) + \langle d, x - \overline{x} \rangle - L \left\|x - \overline{x}\right\|^2 \ \forall x \in U_\varepsilon(\overline{x}) \ \},$$

and the limiting subdifferential of F at $\overline{x}$ is

$$\partial F(\overline{x}) = \left\{ d \in \mathbb{R}^n : \ d = \lim_{k \to \infty} d^k, \ d^k \in \partial^\pi F(x^k), \ x^k \overset{k \to \infty}{\longrightarrow} \overline{x} \right\}$$

These definitions come from [16]. It follows directly from the definitions that $\partial^\pi F(\overline{x}) \subseteq \partial F(\overline{x})$. Furthermore, if the function F is strictly differentiable, the subdifferential can be characterized via $\partial F(\overline{x}) = \{\nabla F(\overline{x})\}$.

Definition 2.2. The directional derivative of $F : \mathbb{R}^n \to \mathbb{R}$ at $\overline{x}$ in a direction $h \in \mathbb{R}^n$ is given by

$$F'(\overline{x}, h) = \lim_{t \downarrow 0} \frac{F(\overline{x} + th) - F(\overline{x})}{t}.$$

If F is differentiable at $\overline{x}$, then the directional derivative at $\overline{x}$ exists for every direction $h \in \mathbb{R}^n$. Furthermore, it holds that $F'(\overline{x}, h) = \nabla F(\overline{x})^\top h$.

Definition 2.3. The proximal normal cone of a closed set $M \subseteq \mathbb{R}^n$ at $\overline{x} \in M$ is defined by

$$N_M^\pi(\overline{x}) = \left\{ d \in \mathbb{R}^n : \exists\, L > 0 \text{ with } \langle d, x - \overline{x} \rangle \le L \|x - \overline{x}\|^2 \ \forall x \in M \right\},$$

the limiting normal cone of M at $\overline{x} \in M$ is its upper limit in the sense of Kuratowski-Painlevé:

$$N_M(\overline{x}) = \left\{ \lim_{k \to \infty} d^k \in \mathbb{R}^n : d^k \in N_M^\pi(x^k),\ x^k \overset{k\to\infty}{\longrightarrow} \overline{x} \right\}.$$

We complete this section with some properties of point-to-set-mappings.

Definition 2.4. A point-to-set-mapping $\Psi : \mathbb{R}^n \to 2^{\mathbb{R}^m}$ is lower semicontinuous at $\overline{x}$ if for every open set Ω with $\Omega \cap \Psi(\overline{x}) \neq \emptyset$, there exists $\varepsilon = \varepsilon(\Omega) > 0$ such that

$$\Psi(\overline{x}) \cap \Omega \neq \emptyset \ \ \forall x \in U_\varepsilon(\overline{x}).$$

Definition 2.5. A point-to-set-mapping $\Psi : \mathbb{R}^n \to 2^{\mathbb{R}^m}$ is called locally upper-Lipschitzian at $\overline{x}$ with modulus L if there exists $\varepsilon > 0$ such that

$$\Psi(x) \subseteq \Psi(\overline{x}) + L \|x - \overline{x}\| \mathbb{B}$$

holds for every $x \in U_\varepsilon(\overline{x})$.

Finally, we need to introduce a suitable substitute for the definition of a derivative for point-to-set-mappings which goes back to Mordukhovich [16].

Definition 2.6. Let $\Psi : \mathbb{R}^n \to 2^{\mathbb{R}^m}$ be a point-to-set-mapping and let $(\overline{x}, \overline{y}) \in \text{cl gph } \Psi$. Then, the coderivative $D^*\Psi : \mathbb{R}^m \to 2^{\mathbb{R}^n}$ of Ψ at $(\overline{x}, \overline{y})$ is defined by

$$D^*\Psi(\overline{x}, \overline{y})(y^*) = \{x^* \in \mathbb{R}^n : (x^*, -y^*) \in N_{\text{gph } \Psi}(\overline{x}, \overline{y}))\}.$$

For a detailed analysis and properties of the aforementioned concepts, the reader is referred to [16] and [18].

3 Stationarity Concepts

At the beginning of this section, we formulate optimality concepts for the KKT reformulation (1.2). There exist different stationarity concepts for mathematical problems with equilibrium constraints, see for example [6]. The special structure of the bilevel programming problem, especially the lower level stationary conditions, make a direct application difficult. Great contribution to the refinement of the conditions was done for example in [4]. We want to confine our studies to two approaches. M-stationarity was first introduced in [23].

Definition 3.1. Consider (1.2). A feasible point $(\overline{x},\overline{y}) \in \text{gph}\,\Psi$ with $G(\overline{x}) \leq 0$ is called M-stationary if there exist multipliers $(\alpha,\beta,\gamma,\overline{\lambda}) \in \mathbb{R}^k \times \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^p$ such that the following conditions hold

$$\nabla_x F(\overline{x},\overline{y}) + \nabla G(\overline{x})^\top \alpha + \nabla_x g(\overline{x},\overline{y})^\top \beta + \nabla_x \mathcal{L}(\overline{x},\overline{y},\overline{\lambda})^\top \gamma = 0 \tag{3.1}$$

$$\nabla_y F(\overline{x},\overline{y}) + \nabla_y g(\overline{x},\overline{y})^\top \beta + \nabla_y \mathcal{L}(\overline{x},\overline{y},\overline{\lambda})^\top \gamma = 0 \tag{3.2}$$

$$\alpha \geq 0,\ \alpha^\top G(\overline{x}) = 0 \tag{3.3}$$

$$\overline{\lambda} \geq 0,\ \overline{\lambda}^\top g(\overline{x},\overline{y}) = 0 \tag{3.4}$$

$$\beta_i = 0 \quad \forall i: \quad i \in \eta \tag{3.5}$$

$$\nabla_y g_i(\overline{x},\overline{y})\gamma = 0 \quad \forall i: \quad i \in \nu \tag{3.6}$$

$$\beta_i > 0,\ \nabla_y g_i(\overline{x},\overline{y})\gamma > 0 \ \vee\ \beta_i(\nabla_y g_i(\overline{x},\overline{y})\gamma) = 0 \quad \forall i: \quad i \in \mu. \tag{3.7}$$

Conditions (3.1) - (3.7) are called M-stationary conditions for problem (1.2).

The second concept we consider is an intensification of Definition 3.1.

Definition 3.2. Consider (1.2). A feasible point $(\overline{x},\overline{y}) \in \text{gph}\,\Psi$ with $G(\overline{x}) \leq 0$ is called S-stationary if there exist multipliers $(\alpha,\beta,\gamma,\overline{\lambda}) \in \mathbb{R}^k \times \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^p$ such that (3.1) - (3.6) and additionally

$$\beta_i \geq 0,\ \nabla_y g_i(\overline{x},\overline{y})\gamma \geq 0 \quad \forall i: i \in \mu. \tag{3.8}$$

holds. These conditions are called S-stationary conditions for problem (1.2).

It follows directly from the definitions that every S-stationary point of (1.2) also fulfills the M-stationary conditions, i.e. every S-multiplier is also an M-multiplier.

In [5], another approach is used to derive necessary optimality conditions for the KKT problem (1.2). There, the basis is the Guignard constraint qualification [17] which demands the equivalence between the dual of the linearized tangent cone and the negative of the normal cone. The Guignard constraint qualification corresponds to the dualized Abadie constraint qualification [17]. The following necessary optimality conditions for problem (1.2) are achieved for $\sigma \in \mathbb{R}_+ := \{t \in \mathbb{R} : t \geq 0\}$:

$$\begin{aligned}
\nabla_x F(\overline{x},\overline{y}) + \nabla G(\overline{x})^\top \alpha + \nabla_x g(\overline{x},\overline{y})^\top (\beta - \sigma\overline{\lambda}) + \nabla_x \mathcal{L}(\overline{\lambda},\overline{x},\overline{y})^\top \gamma &= 0 \\
\nabla_y F(\overline{x},\overline{y}) + \nabla_y g(\overline{x},\overline{y})^\top (\beta - \sigma\overline{\lambda}) + \nabla_y \mathcal{L}(\overline{\lambda},\overline{x},\overline{y})^\top \gamma &= 0 \\
\alpha \geq 0,\ \alpha^\top G(\overline{x}) &= 0 \\
\beta \geq 0,\ \beta^\top g(\overline{x},\overline{y}) &= 0 \\
\overline{\lambda} \geq 0,\ \overline{\lambda}^\top g(\overline{x},\overline{y}) &= 0 \\
\nabla_y g(\overline{x},\overline{y})\gamma - \sigma g(\overline{x},\overline{y}) \geq 0,\ \overline{\lambda}^\top (\nabla_y g(\overline{x},\overline{y})\gamma) &= 0.
\end{aligned} \tag{3.9}$$

The explicit derivation of the conditions (3.9) can be found in [5]. Our aim now is to establish a connection between the conditions (3.9) and S-stationarity. Related relations between S-stationarity conditions for an MPEC and KKT conditions of a nonlinear optimization problem can be found in [7].

First, let $\left(\overline{\alpha}, \overline{\beta}, \overline{\gamma}, \overline{\sigma}\right)$ be the multiplier of a point $(\overline{x}, \overline{y})$ fulfilling the conditions (3.9) with the corresponding multiplier $\overline{\lambda}$. We show that $(\overline{\alpha}, \overline{\beta} - \overline{\sigma}\overline{\lambda}, \overline{\gamma}, \overline{\lambda})$ is an S-stationary multiplier for $(\overline{x}, \overline{y})$. Obviously, conditions (3.1)-(3.4) follow directly from (3.9). Condition (3.5) is also valid since $\overline{\beta}_i - \overline{\sigma}\overline{\lambda}_i = \overline{\beta}_i = 0$ for $i \in \eta$ because of the complementarity conditions $\overline{\beta}^\top g(\overline{x}, \overline{y}) = 0$ in (3.9). We get condition (3.6) from $\overline{\lambda}^\top (\nabla_y g(\overline{x}, \overline{y})\overline{\gamma}) = 0$ since $\overline{\lambda}_i > 0$ for $i \in \nu$. Finally, we verify condition (3.8). For the index set μ, it holds that $\overline{\beta}_i - \overline{\sigma}\overline{\lambda}_i = \overline{\beta}_i \geq 0$ as demanded in (3.9). From $\nabla_y g(\overline{x}, \overline{y})\overline{\gamma} - \overline{\sigma} g(\overline{x}, \overline{y}) \geq 0$ we get $\nabla_y g_i(\overline{x}, \overline{y})\overline{\gamma} \geq 0$ since $g_i(\overline{x}, \overline{y}) = 0$ for $i \in \mu$. Hence, $(\overline{x}, \overline{y})$ is an S-stationary point of (1.2) with the multiplier $(\overline{\alpha}, \overline{\beta} - \overline{\sigma}\overline{\lambda}, \overline{\gamma})$.

For the reverse direction, i.e. in order to derive the stationarity conditions (3.9) from the S-stationary conditions in Definition 3.2, some conditions on the value of σ need to be imposed. They are formulated in the following theorem.

Theorem 3.3. *Let $(\overline{x}, \overline{y})$ be an S-stationary point of problem (1.2) with the corresponding multiplier $\left(\alpha, \beta, \gamma, \overline{\lambda}\right) \in \mathbb{R}^{k+p+m+p}$ and the index sets η, ν, μ. Then, there exists $\sigma \in \mathbb{R}$ with*

$$\beta_i + \sigma\overline{\lambda}_i \quad \geq 0 \quad \forall i \in \nu \tag{3.10}$$

$$\nabla_y g_i(\overline{x}, \overline{y})\gamma - \sigma g_i(\overline{x}, \overline{y}) \quad \geq 0 \quad \forall i \in \eta. \tag{3.11}$$

Then $(\overline{x}, \overline{y})$ fulfills (3.9) with the corresponding multipliers $\left(\alpha, \beta + \sigma\overline{\lambda}, \gamma, \sigma\right)$.

Proof. We only need to ensure the validity of

$$\begin{aligned} \beta + \sigma\overline{\lambda} &\geq 0 \\ \left(\beta + \sigma\overline{\lambda}\right)^\top g(\overline{x}, \overline{y}) &= 0 \\ \nabla_y g(\overline{x}, \overline{y})\gamma &\geq \sigma g(\overline{x}, \overline{y}) \\ \overline{\lambda}^\top \nabla_y g(\overline{x}, \overline{y})\gamma &= 0 \end{aligned}$$

in the conditions (3.9) for all three index sets η, ν, μ. The complementarity conditions follow easily by the definition of η, ν, μ. For the nonnegativity of

$$\beta_i + \sigma\overline{\lambda}_i \ \forall i \in \nu,$$

we only have $\overline{\lambda}_i > 0$ and further need to assume (3.10). Analogously, for $i \in \eta$ we only know that $g_i(\overline{x}, \overline{y}) < 0$ which demands assumption (3.11).

A correct selection of the multiplier σ is ensured by choosing

$$\sigma \geq \max\left\{0, \max_{i \in \nu}\left\{-\frac{\beta_i}{\overline{\lambda}_i}\right\}; \max_{i \in \eta}\left\{\frac{\nabla_y g_i(\overline{x}, \overline{y})\gamma}{g_i(\overline{x}, \overline{y})}\right\}\right\} \tag{3.12}$$

because the other conditions in (3.9) are trivially fulfilled. Due to the assumptions on the occurring variables in (3.12), the maximum is well-defined, hence the existence of the multiplier is ensured. □

Note that conditions (3.10) and (3.11) do not pose additional assumptions since the right-hand side of inequality (3.12) is finite if $(\overline{x},\overline{y})$ is an S-stationary point.

In order to formulate optimality conditions for the KKT reformulation (1.2), suitable constraint qualifications need to be introduced. Due to the complementarity terms in (1.2), it is easy to prove that the usual constraint qualifications used for nonlinear problems like the MFCQ and therefore also the linear independence constraint qualification (LICQ) do not hold at any feasible point of (1.2), see for example [24].

A constraint qualification for M-stationary points is the no nonzero abnormal multiplier constraint qualification (NNAMCQ). For the definition and proof of the lemma regarding the standard MPEC version see for example [22]. For nonlinear programs, NNAMCQ is equivalent to the MFCQ.

Definition 3.4. At a feasible point $(\overline{x},\overline{y},\overline{\lambda})$ of (1.2), NNAMCQ is said to be fulfilled if there does not exist any nonzero vector (α,β,γ) satisfying

$$\begin{aligned}
\nabla G(\overline{x})^\top\alpha+\nabla_x g(\overline{x},\overline{y})^\top\beta+\nabla_x\mathcal{L}(\overline{x},\overline{y},\overline{\lambda})^\top\gamma &= 0\\
\nabla_y g(\overline{x},\overline{y})^\top\beta+\nabla_y\mathcal{L}(\overline{x},\overline{y},\overline{\lambda})^\top\gamma &= 0\\
\mathcal{L}(\overline{x},\overline{y},\overline{\lambda}) &= 0\\
\alpha_i &\geq 0 \quad \forall i: \quad i\in\theta\\
\beta_i &= 0 \quad \forall i: \quad i\in\eta\\
\nabla_y g_i(\overline{x},\overline{y})\gamma &= 0 \quad \forall i: \quad i\in\nu\\
\text{either } \beta_i>0, \nabla_y g_i(\overline{x},\overline{y})\gamma &> 0\\
\text{or } \beta_i(\nabla_y g_i(\overline{x},\overline{y})\gamma) &= 0 \quad \forall i: \quad i\in\mu.
\end{aligned}$$

Lemma 3.5. *Let $(\overline{x},\overline{y},\overline{\lambda})$ be a local optimal solution of* (1.2) *at which NNAMCQ is fulfilled. Then, there exist multipliers (α,β,γ) such that the M-stationarity conditions* (3.1) - (3.7) *are fulfilled at $(\overline{x},\overline{y})$.*

The proof of this Lemma is similar to the respective proof in [22]. We add it for the sake of completeness and since the considered problems here and in [22] are different.

Proof. Let $(\overline{x},\overline{y},\overline{\lambda})$ be locally optimal for (1.2). This is equivalent to $(\overline{x},\overline{y},\overline{\lambda},\overline{a},\overline{b})$ with $\overline{a}=-g(\overline{x},\overline{y})$, $\overline{b}=\overline{\lambda}$ being a local optimal solution of the problem

$$\begin{aligned}
&\min_{x,y,\lambda,a,b} F(x,y)\\
&\text{s.t.} \begin{cases} G(x)\leq 0\\ g(x,y)+a=0,\ -\lambda+b=0\\ \mathcal{L}(x,y,\lambda)=0\\ (a,b)\in\Omega \end{cases}
\end{aligned} \tag{3.13}$$

which is equivalent to (1.2) for $\Omega=\left\{(a,b)\in\mathbb{R}^p\times\mathbb{R}^p:\ a\geq 0,\ b\geq 0,\ a^\top b=0\right\}$. Since $(\overline{x},\overline{y},\overline{\lambda},\overline{a},\overline{b})$ is locally optimal for (3.13), there exist multipliers $\chi\geq 0$, $(\alpha,\beta,\gamma,\xi)\neq(0,0,0,0)$

and $(a^*, b^*) \in N_\Omega(\overline{a}, \overline{b})$ with

$$\begin{aligned}
0 = \chi &\begin{pmatrix} \nabla_x F(\overline{x},\overline{y})^\top \\ \nabla_y F(\overline{x},\overline{y})^\top \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \nabla G(\overline{x})^\top \alpha \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \nabla_x g(\overline{x},\overline{y})^\top \beta \\ \nabla_y g(\overline{x},\overline{y})^\top \beta \\ 0 \\ \beta \\ 0 \end{pmatrix} \\
&+ \begin{pmatrix} \nabla_x \mathcal{L}(\overline{x},\overline{y},\overline{\lambda})^\top \gamma \\ \nabla_y \mathcal{L}(\overline{x},\overline{y},\overline{\lambda})^\top \gamma \\ \nabla_y g(\overline{x},\overline{y})\gamma \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ -\xi \\ 0 \\ \xi \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ a^* \\ b^* \end{pmatrix} \\
&\alpha_i \geq 0 \quad \forall i : i \in \theta \\
&\alpha^\top G(\overline{x}) = 0
\end{aligned} \tag{3.14}$$

as follows from [16]. It holds that $(a^*, b^*) \in N_\Omega(\overline{a}, \overline{b})$ with

$$N_\Omega(a,b) = \{(a^*,b^*) : \quad \begin{array}{ll} a^* = 0 & \text{if } a > 0 \\ b^* = 0 & \text{if } b > 0 \\ a^*, b^* > 0 \text{ or } {a^*}^\top b^* = 0 & \text{if } \ a = b = 0\} \,. \end{array} \tag{3.15}$$

NNAMCQ means that w.l.o.g., χ can be chosen as 1. Conditions (3.1) and (3.2) of the definition of M-stationarity follow directly from (3.14). Furthermore, from (3.14) we also conclude $\nabla_y g(\overline{x},\overline{y})\gamma - \xi = 0$, hence $\xi = \nabla_y g(\overline{x},\overline{y})\gamma$. Taking into account that $a > 0$ whenever $g(x,y) < 0$, which follows from problem (3.13), as well as considering the formulation of the normal cone (3.15), we deduce from $\beta + a^* = 0$ that $\beta_i = 0$ for $i \in \eta$. Following analogous argumentations for the sets ν and μ, we get the relations (3.6) and (3.7). The proof is complete. $\square$

A constraint qualification for S-stationarity is an adjustment of the MPEC-LICQ, for the standard version see [14], see also [21].

Definition 3.6. Let $(\overline{x},\overline{y},\overline{\lambda})$ be feasible for (1.2). MPEC-LICQ holds at the point $(\overline{x},\overline{y},\overline{\lambda})$ if the matrix

$$\begin{pmatrix} \nabla G_\theta(\overline{x}) & 0 & 0 \\ \nabla_x g_{\nu\cup\mu}(\overline{x},\overline{y}) & \nabla_y g_{\nu\cup\mu}(\overline{x},\overline{y}) & 0 \\ \nabla_x \mathcal{L}(\overline{x},\overline{y},\overline{\lambda}) & \nabla_y \mathcal{L}(\overline{x},\overline{y},\overline{\lambda}) & \nabla_y g_\nu^\top(\overline{x},\overline{y}) \end{pmatrix}$$

has full row rank.

Scholtes and Stöhr showed in [21] that MPEC-LICQ holds generically for the problem (1.2).

It is commonly known that MPEC-LICQ is equivalent to the usual nonlinear LICQ for the relaxed problem

$$\min_{x,y,\lambda} F(x,y)$$
$$\text{s.t.} \begin{cases} G(x) \le 0 \\ g(x,y) \le 0,\ \lambda \ge 0, \\ \mathcal{L}(x,y,\lambda) := \nabla_y f(x,y) + \nabla_y g(x,y)^\top \lambda = 0 \end{cases} \tag{3.16}$$

since the complementarity terms $\lambda^\top g(x,y) = 0$ are dropped in this problem; for an explicit proof see [14]. Under MPEC-LICQ, the multipliers corresponding to an S- (hence M-) stationary point are uniquely defined which implies that the two stationarity definitions coincide. Furthermore, this means that for every branch of (1.3) arising from a partition (A,B) of μ, the multiplier is unique, and therefore the multiplier is independent of the partition (A,B).

The following lemma gives some relations between local optimal solutions and S-stationary points of problem (1.2) under MPEC-LICQ. It is an application of results in [20].

Lemma 3.7. *Let $(\overline{x},\overline{y},\overline{\lambda})$ be a local optimal solution of* (3.16)*, and let LICQ hold at this point. Then, there exist multipliers (α,β,γ) such that $(\overline{x},\overline{y})$ is an S-stationary point for* (1.2) *with multipliers $(\alpha,\beta,\gamma,\overline{\lambda})$.*

Proof. The optimality conditions for the relaxed problem (3.16) with respect to $(\overline{x},\overline{y},\overline{\lambda})$ are

$$0 = \begin{pmatrix} \nabla_x F(\overline{x},\overline{y})^\top \\ \nabla_y F(\overline{x},\overline{y})^\top \\ 0 \end{pmatrix} + \begin{pmatrix} \nabla G(\overline{x})^\top \alpha \\ 0 \\ 0 \end{pmatrix}$$
$$+ \begin{pmatrix} \nabla_x g(\overline{x},\overline{y})^\top \beta \\ \nabla_y g(\overline{x},\overline{y})^\top \beta \\ 0 \end{pmatrix} + \begin{pmatrix} \nabla_x \mathcal{L}(\overline{x},\overline{y},\overline{\lambda})^\top \gamma \\ \nabla_y \mathcal{L}(\overline{x},\overline{y},\overline{\lambda})^\top \gamma \\ \nabla_y g(\overline{x},\overline{y})\gamma \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ \xi \end{pmatrix} \tag{3.17}$$

$$\alpha_i \ge 0 \qquad \forall i:\ i \in \theta \tag{3.18}$$
$$\beta_i \ge 0 \qquad \forall i:\ i \in \nu \cup \mu \tag{3.19}$$
$$\xi_i \ge 0 \qquad \forall i:\ i \in \eta \cup \mu \tag{3.20}$$
$$\alpha^\top G(\overline{x}) = 0 \tag{3.21}$$
$$\beta^\top g(\overline{x},\overline{y}) = 0 \tag{3.22}$$
$$\xi^\top \lambda = 0. \tag{3.23}$$

From (3.22) we deduce that $\beta_i = 0$ for $i \in \eta$ which is condition (3.5) of S-stationarity. Since $\xi = \nabla_y g(\overline{x},\overline{y})\gamma$ which we get from (3.17), we can use (3.23) in order to derive $\nabla_y g_i(\overline{x},\overline{y})\gamma = 0$ for all $i \in \nu$ which equals the S-stationarity condition (3.6). Finally, the nonnegativity (3.8) of β_i and $\nabla_y g_i(\overline{x},\overline{y})\gamma$ for $i \in \mu$ follows from (3.19) and (3.20). The proof is complete. □

4 Strongly Stable M- and S-Stationary Points

Our aim is to provide conditions for strong stability of the introduced stationary points.

Definition 4.1. An M-stationary (S-stationary) point $(\overline{x},\overline{y})$ of (1.2) is called strongly stable in the sense of Kojima [13] if for some $r > 0$ and each $\varepsilon \in (0,r]$ there exists a $\delta = \delta(\varepsilon) > 0$

such that for every $(\tilde{F}, \tilde{f}, \tilde{G}, \tilde{g}) \in C^2$ with

$$\|(F-\tilde{F}, f-\tilde{f}, G-\tilde{G}, g-\tilde{g})\|^{C^2}_{U_r(\overline{x},\overline{y})} \leq \delta,$$

there exists $(\tilde{x}, \tilde{y})$ which is M-stationary (S-stationary) for (1.2) in $U_\varepsilon(\overline{x},\overline{y})$ and unique in $U_r(\overline{x},\overline{y})$.

Turning our attention to problem (1.2) and the corresponding index sets μ, ν, η, potential problems arise whenever the degenerate index set μ is nonempty. Having a closer look at the stationarity conditions, it becomes obvious that a deeper classification is possible. Therefore, considering an M-stationary point $(\overline{x},\overline{y})$ and the corresponding multipliers $(\alpha, \beta, \gamma, \overline{\lambda})$ of (1.2), we introduce for $\zeta_i := \nabla_y g_i(\overline{x},\overline{y})\gamma$ the following partition of the index set μ:

$$\begin{aligned}
p &:= \{i \in \mu :\, \beta_i > 0,\, \zeta_i > 0\}\\
q &:= \{i \in \mu :\, \beta_i > 0,\, \zeta_i = 0\}\\
r &:= \{i \in \mu :\, \beta_i = 0,\, \zeta_i > 0\}\\
s &:= \{i \in \mu :\, \beta_i < 0,\, \zeta_i = 0\}\\
t &:= \{i \in \mu :\, \beta_i = 0,\, \zeta_i < 0\}\\
v &:= \{i \in \mu :\, \beta_i = 0,\, \zeta_i = 0\}.
\end{aligned} \tag{4.1}$$

Now, the possibly degenerate structure is reduced to the set $v \subseteq \mu$.

Remark 4.2. Regarding S-stationary points, the index sets s and t are by definition empty.

The next lemma provides a characterization of stationary points of the KKT problem (1.2) using the partition (4.1).

Lemma 4.3. *Let $(\overline{x},\overline{y})$ be M-stationary for* (1.2) *with the multiplier $(\overline{\alpha},\overline{\beta},\overline{\gamma},\overline{\lambda})$. Assume that the set v is empty. Let the mapping $\mathcal{H}$ be defined by*

$$\mathcal{H}(x,y,\alpha,\beta,\gamma,\lambda) := \begin{pmatrix}
\nabla_x F(x,y) + \nabla G(x)^\top \alpha + \nabla_x g(x,y)^\top \beta + \nabla_x \mathcal{L}(x,y,\lambda)^\top \gamma\\
\nabla_y F(x,y) + \nabla_y g(x,y)^\top \beta + \nabla_y \mathcal{L}(x,y,\lambda)^\top \gamma\\
\min\{\alpha, -G(x)\}\\
\min\{\lambda, -g(x,y)\}\\
\mathcal{L}(x,y,\lambda)\\
g_\nu(x,y)\\
\lambda_\eta\\
g_p(x,y)\\
\lambda_p\\
g_q(x,y)\\
\min\{\zeta_q, \lambda_q\}\\
\min\{\beta_r, -g_r(x,y)\}\\
\lambda_r\\
g_s(x,y)\\
\zeta_s\\
\beta_t\\
\lambda_t
\end{pmatrix}.$$

Then, in a sufficiently small neighborhood of $(\overline{x}, \overline{y})$*, every M-stationary point* $(\widehat{x}, \widehat{y})$ *of* (1.2) *with multipliers* $(\widehat{\alpha}, \widehat{\beta}, \widehat{\gamma}, \widehat{\lambda})$ *sufficiently close to* $(\overline{\alpha}, \overline{\beta}, \overline{\gamma}, \overline{\lambda})$ *is a root of the mapping* $\mathcal{H}$ *and vice versa.*

Proof. That every M-stationary point is a root of $\mathcal{H}$ follows directly from Definition 3.1 and the partition of the index set μ. The closeness condition is used to guarantee that the index sets ν, μ, and η do not change. For the reverse direction of the Lemma, the validity of M-stationary conditions (3.4) and (3.7) have to be verified for the respective index sets ν, μ, and η. □

The condition that $(\widehat{\alpha}, \widehat{\beta}, \widehat{\gamma}, \widehat{\lambda})$ is sufficiently close to $(\overline{\alpha}, \overline{\beta}, \overline{\gamma}, \overline{\lambda})$ is guaranteed under MPEC-LICQ.

Proposition 4.4. *Let* $(\overline{x}, \overline{y})$ *be an S-stationary point of* (1.2) *for the multipliers* $\left(\overline{\alpha}, \overline{\beta}, \overline{\gamma}, \overline{\lambda}\right)$*. The first direction of Lemma 4.3, i.e.* $\mathcal{H}(x, y, \alpha, \beta, \gamma, \lambda) = 0$ *for every S-stationary point* $(x, y, \alpha, \beta, \gamma, \lambda) \in U(\overline{x}, \overline{y}, \overline{\alpha}, \overline{\beta}, \overline{\gamma}, \overline{\lambda})$*, holds automatically since S-stationarity implies M-stationarity. For the reverse direction, we further need to assume that besides* v*, also the index sets* s, t *are empty since S-stationarity requires the multipliers* β_i, ζ_i *to be nonnegative for every* $i \in \mu$*.*

A standard way to achieve strong stability for nonlinear programs is via LICQ and sufficient conditions of second order. We formulate such conditions in Theorem 5.6 of Section 5. Another approach appears in [12] where regularity of the matrices $\partial\mathcal{H}$ is used. The next theorem is based on the latter principle and provides sufficient conditions for strong stability.

Theorem 4.5. *Let* $(\overline{x}, \overline{y})$ *be an M-stationary point of* (1.2) *with the multipliers* $(\overline{\alpha}, \overline{\beta}, \overline{\gamma}, \overline{\lambda})$*. Consider* $\mathcal{H}(\overline{x}, \overline{y}, \overline{\alpha}, \overline{\beta}, \overline{\gamma}, \overline{\lambda})$ *as introduced above. If* $v = \emptyset$*, all the matrices* $\partial\mathcal{H}(\overline{x}, \overline{y}, \overline{\alpha}, \overline{\beta}, \overline{\gamma}, \overline{\lambda})$ *are regular and MPEC-LICQ holds, then* $(\overline{x}, \overline{y})$ *is a strongly stable M-stationary point of* (1.2)*.*

Proof. With the help of Lemma 4.3, we get $\mathcal{H}(\overline{x}, \overline{y}, \overline{\alpha}, \overline{\beta}, \overline{\gamma}, \overline{\lambda}) = 0$. This allows us to apply Theorem 4.3 and Remark 4.4 of [12] which states the equivalence of non-singularity of all the matrices in the Clarke's subdifferential $\partial\mathcal{H}(\overline{x}, \overline{y}, \overline{\alpha}, \overline{\beta}, \overline{\gamma}, \overline{\lambda})$ and strong stability of $(\overline{x}, \overline{y})$ under LICQ. □

Proposition 4.6. *Under the assumptions of Theorem 4.5, the statement can be applied directly to an S-stationary point of* (1.2) *since every S-stationary point is also M-stationary.*

For the investigation of C-stationarity, a concept that is weaker than M-stationarity, we refer the interested reader to [11]. There, a full characterization of strongly stable C-stationary points is given under similar assumtions.

$\boxed{5}$ Properties of the Nonlinear Program

We turn our attention to the nonlinear program (1.2) and split it into a patchwork of problems (1.3). Then, properties of solutions and of the optimal value functions of these problems can be used to derive properties of the problem (1.2).

For a feasible point $(\overline{x}, \overline{y}, \overline{\lambda})$ of the KKT problem (1.2), we consider one nonlinear problem (1.3) arising from a special partition (A, B) of the degenerate index set μ. In order to study the stability of a locally optimal solution of (1.3), we need to investigate its parameterized optimal value function

$$\varphi_{NLP}(x) := \min_{(y,\lambda)\in\mathbb{R}^m} \{F(x,y) : (y,\lambda) \in M_{NLP}(x)\}$$

with the set of feasible points

$$\begin{aligned} M_{NLP}(x) := \big\{(y,\lambda) \in \mathbb{R}^{m+p} : \ & G(x) \leq 0, \\ & g_i(x,y) = 0,\ \lambda_i \geq 0 \quad i \in \nu \cup A \\ & g_i(x,y) \leq 0,\ \lambda_i = 0 \quad i \in \eta \cup B \\ & \mathcal{L}(x,y,\lambda) = 0\} \end{aligned}$$

and the set of optimal solutions

$$\Psi_{NLP}(x) = \operatorname*{Argmin}_{(y,\lambda)\in M_{NLP}(x)} F(x,y).$$

We get the relation

$$\varphi(x) = \min_{(A,B)} \varphi_{NLP}(x)$$

for the optimal value function of the original problem.

First, we get the following representation for the directional derivative of the optimal value function. The idea for this representation stems from [8] and [19].

Lemma 5.1. *Suppose that there exist $\kappa \in \mathbb{R}$ and a compact set $S \subseteq \mathbb{R}^{m+p}$ with $\varphi_{NLP}(\overline{x}) < \kappa$ and*

$$\{(y,\lambda) \in M_{NLP}(x) : F(x,y) \leq \kappa\} \subseteq S \quad \forall x \in U(\overline{x}). \tag{5.1}$$

Furthermore, let LICQ be fulfilled for every $(y,\lambda) \in \Psi_{NLP}(\overline{x})$. Then,

$$\varphi'_{NLP}(\overline{x},h) = \min_{(y,\lambda)\in\Psi_{NLP}(\overline{x})} \{\nabla_x \mathbb{L}(\overline{x},y,\alpha,\beta,\gamma,\lambda)h\}$$

with the Lagrangian function

$$\mathbb{L}(x,y,\alpha,\beta,\gamma,\lambda) = F(x,y) + \alpha^\top G(x) + \beta^\top g(x,y) + \gamma^\top \mathcal{L}(x,y,\lambda).$$

Proof. Consider the problem (1.3). Under LICQ, the Lagrangian multiplier is defined uniquely. The existence of a solution (y,λ) in a sufficiently small neighborhood of $\overline{x}$ as long as $M_{NLP}(x) \neq \emptyset$ is ensured by the condition (5.1). In order to derive the result from [8], we only need to ensure that Ψ_{NLP} is uniformly compact. This follows from condition (5.1) which ensures the boundedness of the level sets $\{(y,\lambda) \in M_{NLP}(x) : F(x,y) \leq \kappa\}$ together with [16] which ensures the closedness. $\square$

Remark 5.2. Condition (5.1) is called inf-compactness condition. The formula for the directional differentiability in the form used here can also been found in the paper [10].

Remark 5.3. It is possible to presuppose the inf-compactness condition with respect to the KKT problem (1.2), its parameter-dependent feasible set $M(x)$ and optimal value function $\varphi(x)$. This is valid because for every branch, i.e. for every partition (A,B), and for κ and S as introduced in Lemma 5.1 we get $\varphi(\overline{x}) = \varphi_{NLP}(\overline{x}) < \kappa$ and

$$\{(y,\lambda) \in M_{NLP}(x) : F(x,y) \leq \kappa\} \subseteq \{(y,\lambda) \in M(x) : F(x,y) \leq \kappa\} \subseteq S$$

for all $x \in U(\overline{x})$. Analogously, the LICQ condition for the nonlinear program can be substituted by MPEC-LICQ for the KKT reformulation since in a small neighborhood of $(\overline{x}, \overline{y})$ and for every partition (A, B), the feasible set of (1.3) is contained in the one of the original problem (1.2). This means that MPEC-LICQ ensures LICQ for every branch of (1.3).

Example 5.4. We want to examine the bilevel programming problem

$$\begin{array}{ll} \min\limits_{x,y} & F(x,y) = (x-1)^2 \\ \text{s.t.} & y \in \Psi(x) := \operatorname*{Argmin}\limits_{z} \; \{ \, x^2 + z^2 : \, g_1(x,z) = 1 - x + z \leq 0 \, \} \, . \end{array}$$

The solution of the KKT formulation

$$\begin{array}{ll} \min\limits_{x,y,\lambda} & F(x,y) = (x-1)^2 \\ \text{s.t.} & g_1(x,y) = 1 - x + y \leq 0, \, \lambda \geq 0, \, g_1(x,y) \cdot \lambda = 0 \\ & \mathcal{L}(x,y,\lambda) = 2y + \lambda = 0 \end{array}$$

is $(x, y, \lambda) = (1, 0, 0)$. Hence, the index set $\mu = \{1\}$ leads to the two possible partitions $(A, B) = (\{1\}, \emptyset)$ and $(A, B) = (\emptyset, \{1\})$. Therefore, two nonlinear surrogate problems arise. We confine our studies to the following one with $(A, B) = (\{1\}, \emptyset)$, namely

$$\begin{array}{ll} \min\limits_{x,y,\lambda} & F(x,y) = (x-1)^2 \\ \text{s.t.} & g_1(x,y) = 1 - x + y = 0, \, \lambda_1 \geq 0 \\ & \mathcal{L}_1(x,y,\lambda) = 2y + \lambda = 0. \end{array}$$

It can be easily deduced that the only feasible solution of the latter problem is $(x, y, \lambda) = (1, 0, 0)$ and that LICQ holds at this point. Then, the calculation of the directional derivative for $h \in \mathbb{R}$ reads

$$\varphi'_{NLP}(1, h) = h(2(x-1) - \beta).$$

Evaluating the optimality conditions, we get $\beta = 0$ belonging to the constraint $g_1(x, y) = 1 - x + y = 0$, and $\gamma = 0$ which is the multiplier of the constraint $\mathcal{L}(x, y, \lambda) = 2y + \lambda = 0$. Hence, $\varphi'_{NLP}(1, h) = 0$ for every $h \in \mathbb{R}$.

In order to formulate continuity statements for the optimal value function φ_{NLP}, we need to introduce a second order condition.

Definition 5.5. Let $(\overline{x}, \overline{y})$ be S-stationary for (1.2) with the unique multiplier $(\overline{\alpha}, \overline{\beta}, \overline{\gamma}, \overline{\lambda})$. The strong second order sufficient condition (MPEC-SSOSC) holds at $(\overline{x}, \overline{y})$ if

$$d^\top \nabla^2 \mathbb{L}(\overline{x}, \overline{y}, \overline{\alpha}, \overline{\beta}, \overline{\gamma}, \overline{\lambda}) d > 0$$

for every nonzero vector $d = (dx, dy, d\lambda)^\top \in \mathbb{R}^{n+m+p}$ satisfying

$$\begin{aligned} \nabla G_i(\overline{x}) dx &= 0 \quad \forall i : \alpha_i > 0 \\ \nabla_x g_i(\overline{x}, \overline{y}) dx &= 0 \quad \forall i : i \in \nu \cup \mu \\ \nabla_y g_i(\overline{x}, \overline{y}) dy &= 0 \quad \forall i : i \in \nu \cup \mu \\ \nabla_x \mathcal{L}(\overline{x}, \overline{y}, \overline{\lambda}) dx &= 0 \\ \nabla_y \mathcal{L}(\overline{x}, \overline{y}, \overline{\lambda}) dy &= 0 \\ (\nabla_y g_i(\overline{x}, \overline{y}) \gamma) \, d\lambda &= 0 \quad \forall i : i \in \eta \cup \mu \end{aligned}$$

with $\mathbb{L}(x, y, \alpha, \beta, \gamma, \lambda) = F(x, y) + \alpha^\top G(x) + \beta^\top g(x, y) + \gamma^\top \mathcal{L}(x, y, \lambda)$.

Theorem 5.6. *Let $(\overline{x}, \overline{y})$ be an S-stationary point of* (1.2) *at which MPEC-LICQ and MPEC-SSOSC are fulfilled. Then, φ_{NLP} is Lipschitz-continuous over $U(\overline{x})$ and there exists a unique, continuous function $(\overline{y}(x), \overline{\lambda}(x))$ of optimal solutions with $(\overline{y}(\overline{x}), \overline{\lambda}(\overline{x})) = (\overline{y}, \overline{\lambda})$.*

Proof. We begin with the proof second part of the theorem. From MPEC-LICQ and MPEC-SSOSC we derive that both LICQ and SSOSC hold with respect to every branch (i.e. every partition of μ) of the nonlinear program (1.3). It follows that $(\overline{x}, \overline{y})$ with $(\overline{y}, \overline{\lambda}) \in \Psi_{NLP}(\overline{x})$ is a unique optimal solution for every partition (A, B). This remains true in a sufficiently small neighborhood, i.e. for every $x \in U(\overline{x})$, there exists a unique solution $(\overline{y}(x), \overline{\lambda}(x))$ with $(\overline{y}, \overline{\lambda}) : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}$ being continuous as follows from the implicit function theorem.
The Lipschitz-continuity of φ_{NLP} can be deduced analogously. □

Since for every partition, the corresponding nonlinear program (1.3) respresents a part of the feasible set of the original problem, the investigation in certain directions is of special interest. Therefore, consider the curve

$$x(t) = \overline{x} + th + o(t) \quad 0 \le t < \varepsilon$$

for some $\varepsilon > 0$. Now we want to prove a Lipschitzian property for the optimal set mapping Ψ of (1.3).

Theorem 5.7. *Suppose that LICQ holds for every $(y, \lambda) \in \Psi_{NLP}(\overline{x})$ and that the inf-compactness condition* (5.1) *is fulfilled. For all $(y, \lambda) \in \Psi_{NLP}(\overline{x})$ with*

$$\varphi'_{NLP}(\overline{x}, h) = \nabla_x \mathbb{L}(\overline{x}, y, \alpha, \beta, \gamma, \lambda)h,$$

let there be a sufficiently small neighborhood $U(y, \lambda)$ such that the active index sets

$$\{i \in \nu \cup A : \lambda'_i = 0\} \cup \{i \in \eta \cup B : g_i(\overline{x}, y') = 0\} \tag{5.2}$$

do not change for all $(y', \lambda') \in U(y, \lambda)$, $(y', \lambda') \in \Psi_{NLP}(\overline{x})$ with Ψ_{NLP} being compact. Then, the optimal set mapping Ψ_{NLP} is locally upper Lipschitzian at $\overline{x}$ along the direction h, i.e. there exists $L > 0$ such that

$$\Psi_{NLP}(x(t)) \subseteq \Psi_{NLP}(\overline{x}) + Lt\mathbb{B}$$

holds.

Proof. First, we need to introduce the critical cone to M_{NLP} at $(\overline{x}, y)$ with $(y, \lambda) \in \Psi_{NLP}(\overline{x})$,

$$\begin{aligned} C(y,\lambda) = \{\ (dy, d\lambda) \in \mathbb{R}^{m+p} :\ & \nabla_y g_i(\overline{x}, y)dy = 0 \text{ if } i \in \nu \cup A \\ & d\lambda_i = 0 \text{ if } i \in \nu \cup A,\ \nabla_y g_i(\overline{x}, y)\gamma > 0, \lambda_i = 0 \\ & d\lambda_i \ge 0 \text{ if } i \in \nu \cup A,\ \nabla_y g_i(\overline{x}, y)\gamma = \lambda_i = 0 \\ & \nabla_y g_i(\overline{x}, y)dy = 0 \text{ if } i \in \eta \cup B,\ \beta_i > 0, g_i(\overline{x}, y) = 0 \\ & \nabla_y g_i(\overline{x}, y)dy \le 0 \text{ if } i \in \eta \cup B,\ \beta_i = g_i(\overline{x}, y) = 0 \\ & d\lambda_i = 0 \text{ if } i \in \eta \cup B \\ & \nabla_y \mathcal{L}(\overline{x}, y, \lambda)dy + \nabla_\lambda \mathcal{L}(\overline{x}, y, \lambda)d\lambda = 0\ \} \end{aligned}$$

and the linearized cone

$$\begin{aligned} T^{\text{lin}}(y,\lambda) = \{\ (dy, d\lambda) \in \mathbb{R}^{m+p} :\ & \nabla_y g_i(\overline{x}, y)dy = 0 \text{ if } i \in \nu \cup A \\ & d\lambda_i \ge 0 \text{ if } i \in \nu \cup A,\ \lambda_i = 0 \\ & d\lambda_i = 0 \text{ if } i \in \eta \cup B \\ & \nabla_y g_i(\overline{x}, y)dy \le 0 \text{ if } i \in \eta \cup B,\ g_i(\overline{x}, y) = 0 \\ & \nabla_y \mathcal{L}(\overline{x}, y, \lambda)dy + \nabla_\lambda \mathcal{L}(\overline{x}, y, \lambda)d\lambda = 0\ \}\,. \end{aligned}$$

It follows readily that

$$\begin{aligned}&\big\{(dy, d\lambda) \in C(y,\lambda) :\ dy^\top \nabla^2_{yy}\mathbb{L}(\overline{x}, y, \lambda)dy + d\lambda^\top \nabla^2_{\lambda\lambda}\mathbb{L}(\overline{x}, y, \lambda)d\lambda = 0\big\}\\ &\subseteq C(y,\lambda) \subseteq T^{\mathrm{lin}}(y,\lambda)\end{aligned}$$

holds. Since LICQ holds for every $(y, \lambda) \in \Psi_{NLP}(\overline{x})$, also the Abadie constraint qualification is fulfilled [17] which means that we get the inclusion

$$T^{\mathrm{lin}}(y,\lambda) \subseteq T_{M_{NLP}(\overline{x})}(y,\lambda)$$

for every $(y,\lambda) \in M_{NLP}(\overline{x})$.
Furthermore, the compactness of Ψ_{NLP} allows the formulation of the coderivative at $(\overline{x}, y, \lambda) \in \operatorname{gph} \Psi_{NLP}$ for every $(y, \lambda) \in M_{NLP}(\overline{x})$, see Definition 2.6. As a result from Theorem 5.6, Ψ_{NLP} is a smooth manifold at (y, λ).
Now, all assumptions are fulfilled in order to apply [19, Theorem 3.1]. □

Remark 5.8. A sufficient condition for the index sets (5.2) being constant is that $\mu = \emptyset$.

Acknowledgement

The authors thank the referees for their very helpful remarks which helped us to improve the presentation of the paper.

References

[1] S. Dempe, *Foundations of Bilevel Programming*, Kluwer Academic Publishers, Dordrecht, 2002

[2] S. Dempe and J. Dutta, Is bilevel programming a special case of a mathematical program with complementarity constraints?, *Math. Program.* 131 (2012) 37–48.

[3] S. Dempe, B.S. Mordukhovich and A.B. Zemkoho, Sensitivity analysis for two-level value functions with applications to bilevel programming, Optimization Online.

[4] S. Dempe and A.B. Zemkoho, On the Karush-Kuhn-Tucker reformulation of the bilevel optimization problem, *Nonlinear Analysis* 75 (2012) 1202–1218.

[5] S. Dempe and A.B. Zemkoho, The bilevel programming problem: reformulations, constraint qualifications and optimality conditions, *Math. Program.*, to appear.

[6] M. Flegel and C. Kanzow, On M-stationary points for mathematical programs with equilibrium constraints, *Journal of Mathematical Analysis and Applications* 310 (2005) 286–302.

[7] R. Fletcher, S. Leyffer, D. Ralph and S. Scholtes, Local convergence of SQP methods for mathematical programs with equilibrium constraints, *SIAM J. Optim.* 17 (2006) 259–286.

[8] J. Gauvin and J.W. Tolle, Differential properties of the marginal function in mathematical programming, *Mathematical Programming Study* 19 (1977) 101–119.

[9] J. Guddat, H. Jongen and J. Rückmann, On stability and stationary points in nonlinear optimization, *J. Austral. Math. Soc. Ser. B* 28 (1986) 36–56.

[10] X. Hu and D. Ralph, A note on the sensitivity of value functions of mathematical programs with complementarity constraints, *Mathematical Programming* 93 (2002) 265–279.

[11] H.Th. Jongen, V. Shikhman and S. Steffensen, Characterization of strong stability for C-stationary points in MPCC, *Math. Program.* 132 (2012) 295–308.

[12] D. Klatte and K. Tammer, Strong stability of stationary solutions and Karush-Kuhn-Tucker points in nonlinear optimization, *Ann. Oper. Res.* 27 (1990) 285–308.

[13] M. Kojima, Strongly stable stationary solutions in nonlinear programs, in *Analysis and Computation of Fixed Points*, S.M. Robinson (ed.), Academic Press, New York, 1980, pp. 93–138.

[14] Z. Luo, J. Pang and D. Ralph, *Mathematical Programs with Equilibrium Constraints*, Cambridge University Press, Cambridge, 1996

[15] B.S. Mordukhovich, Maximum principle in problems of time optimal control with non-smooth constraints, *J. Appl. Math. Mech.* 40 (1976) 960–969.

[16] B.S. Mordukhovich, *Variational Analysis and Generalized Differentiation I: Basic Theory*, Springer, Berlin, 2006

[17] D.W. Peterson, A review of constraint qualifications in finite-dimensional spaces, *SIAM Review* 15 (1973) 639–654.

[18] R.T. Rockafellar and R.J.-B. Wets, *Variational Analysis*, Springer-Verlag, Berlin Heidelberg, 1998

[19] A. Shapiro, Perturbation theory of nonlinear programs when the set of optimal solutions is not a singleton, *App. Math. Optim.* 18 (1988) 215–229.

[20] H. Scheel and S. Scholtes, Mathematical programs with equilibrium constraints: stationarity, optimality, and sensitivity, *Mathematics of Operations Research* 25 (2000) 1–22.

[21] S. Scholtes, M. Stöhr, How stringent is the linear independence assumption for mathematical programs with complementarity constraints?, *Math. Oper. Res.* 26 (2001) 851–863.

[22] J.J. Ye, Necessary and sufficient optimality conditions for mathematical programs with equilibrium constraints, *Journal of Mathematical Analysis and Applications* 307 (2005) 350–369.

[23] J.J. Ye and X.Y. Ye, Necessary optimality conditions for optimization problems with variational inequality constraints, *Mathematical Operations and Research* 22 (1997) 977–997.

[24] J.J. Ye and D.L. Zhu, Optimality conditions for bilevel programming problems, *Optimization* 33 (1995) 9–27.

Manuscript received 7 March 2012
revised 5 November 2012
accepted for publication 12 November 2012

S. Dempe
TU Bergakademie Freiberg, Institute of Numerical Mathematics and Optimization, D-09596 Freiberg
E-mail address: dempe@math.tu-freiberg.de

S. Franke
TU Bergakademie Freiberg, Institute of Numerical Mathematics and Optimization, D-09596 Freiberg
E-mail address: franke@math.tu-freiberg.de