Optimization
A Journal of Mathematical Programming and Operations Research

ISSN: 0233-1934 (Print) 1029-4945 (Online) Journal homepage: https://www.tandfonline.com/loi/gopt20

Solution of bilevel optimization problems using the KKT approach

S. Dempe & S. Franke

OPTIMIZATION
https://doi.org/10.1080/02331934.2019.1581192

Solution of bilevel optimization problems using the KKT approach

S. Dempe and S. Franke

TU Bergakademie Freiberg, Freiberg, Germany

ABSTRACT
Using the Karush–Kuhn–Tucker conditions for the convex lower level problem, the bilevel optimization problem is transformed into a single-level optimization problem (a mathematical program with complementarity constraints). A regularization approach for the latter problem is formulated which can be used to solve the bilevel optimization problem. This is verified if global or local optimal solutions of the auxiliary problems are computed. Stationary solutions of the auxiliary problems converge to C-stationary solutions of the mathematical program with complementarity constraints.

ARTICLE HISTORY
Received 2 February 2017
Accepted 1 February 2019

KEYWORDS
Bilevel optimization; Karush–Kuhn–Tucker transformation; mathematical program with complementarity constraints

2010 MATHEMATICS SUBJECT CLASSIFICATIONS
90C26; 91A65; 65K05

1. Introduction

Bilevel optimization problems model the situation of minimizing an objective function subject to the graph of the solution set mapping of a second, parametric optimization problem. In order to formulate it, first consider the latter problem, called the *lower level problem* from now on:

$$\min_{y} \{f(x,y) : g(x,y) \leq 0\}, \tag{1}$$

where $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$ are sufficiently smooth functions and $y \mapsto f(x,y)$, $y \mapsto g_i(x,y)$, $i = 1,\ldots,p$ are convex for each fixed x. For fixed x,

assumes lower lvl problem is cvx

$$\Psi(x) := Argmin\{f(x,y) : g(x,y) \leq 0\}$$

denotes the solution set mapping of the lower level problem and

$$\mathbf{gph}\ \Psi := \{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m : y \in \Psi(x)\}$$

its graph. Then, the *bilevel optimization problem* (or upper level problem) is defined as

$$\min_{x,y} \{F(x,y) : G(x) \leq 0,\ (x,y) \in \mathbf{gph}\ \Psi\}. \tag{2}$$

CONTACT S. Dempe dempe@math.tu-freiberg.de

Here, $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $G : \mathbb{R}^n \to \mathbb{R}^q$ are assumed to be sufficiently smooth functions. Note that we use the optimistic approach.

Bilevel optimization problems have been the topic of the monographs of Bard [1], Dempe [2] and Dempe et al. [3]. They have a large number of applications, see e.g. the bibliography [4]. Even under the assumptions above, they are nonconvex, nonsmooth optimization problems. To solve bilevel optimization problems, they need to be transformed into a single-level optimization problem. In order to achieve this, different approaches are possible, see [3].

The use of the optimal value function

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

of the lower level problem transforms (1), (2) into an equivalent, nonsmooth and nonconvex optimization problem:

$$\min_{x,y} \{F(x,y) : G(x) \le 0,\ f(x,y) \le \varphi(x),\ g(x,y) \le 0\}.$$

Necessary optimality conditions related to this approach can be found in the article [5] of Ye and Zhu or [6], whereas solution algorithms using this approach are the topic of the articles [7, 8] of Dempe and Franke.

Since the lower level problem (1) is assumed to be a convex optimization problem, it can be replaced by its necessary and sufficient optimality condition

$$0 \in \nabla_y f(x,y) + N_{T(x)}(y),$$

where $T(x) := \{y \in \mathbb{R}^m : g(x,y) \le 0\}$ denotes the feasible set of problem (1) for fixed x and $N_{T(x)}(y)$ is the normal cone to $T(x)$ at y in the sense of convex analysis, see the monograph [9] of Rockafellar and Wets. This results in the equivalent nondifferential optimization problem

$$\min_{x,y} \{F(x,y) : G(x) \le 0,\ 0 \in \nabla_y f(x,y) + N_{T(x)}(y)\}.$$

Note that $N_{T(x)}(y) = \emptyset$ if $y \notin T(x)$. Necessary optimality conditions using this primal Karush–Kuhn–Tucker (KKT for short) transformation can be found in Zemkoho's PhD thesis [10].

The KKT conditions for the lower level problem (1) have often been used to transform the bilevel optimization problem. They are necessary and sufficient optimality conditions for (1) provided that some regularity condition is satisfied for this problem. Let the Mangasarian–Fromovitz constraint qualification (MFCQ) be satisfied for (1) at each point $(\bar{x},\bar{y}) \in \mathbf{gph}\ T$ with $G(\bar{x}) \le 0$:

$$\exists\, d \in \mathbb{R}^m : \nabla_y g_i(\bar{x},\bar{y})d < 0\ \forall\, i \in \{j \in \{1,2,\ldots,p\} : g_j(\bar{x},\bar{y}) = 0\}.$$

Then, (1), (2) is transformed into

$$
\begin{aligned}
F(x,y) &\to \min_{x,y,u} \\
G(x) &\leq 0, \\
\nabla_y L(x,y,u) &= 0, \\
g(x,y) &\leq 0, \\
u &\geq 0, \\
u^\top g(x,y) &= 0
\end{aligned}
\tag{3}
$$

with $L(x,y,u) = f(x,y) + u^\top g(x,y)$. Mirrlees [11] has shown that this approach cannot be used if the lower level problem (1) is not a convex optimization problem, see also [8]. The MFCQ is not generically satisfied for the lower level problem at an optimal solution of the bilevel optimization problem, but replacing (1) with the Fritz–John conditions as in Allende and Still's article [12] is also not an equivalent transformation, see [8]. It has been shown that problem (3) is equivalent to (1), (2) if global optimal solutions are investigated. But, as problem (3) is a nonconvex optimization problem, solution algorithms solving this problem usually compute stationary or local optimal solutions. In that case, both problems are not equivalent, as the following statement by Dempe and Dutta [13] shows.

Theorem 1.1 (Dempe and Dutta [13]). *Consider the bilevel optimization problem* (1), (2), *and let MFCQ be satisfied at all points* $(x,y) \in \mathbf{gph}\, T$ *with* $G(x) \leq 0$. *Then,* $(\bar{x},\bar{y})$ *is a local optimal solution of* (1), (2) *if and only if* $(\bar{x},\bar{y},\bar{u})$ *is a local optimal solution of problem* (3) *for all*

$$\bar{u} \in \Lambda(\bar{x},\bar{y}) := \{u \in \mathbb{R}^p : \nabla_y L(\bar{x},\bar{y},u) = 0,\ u \geq 0,\ u^\top g(\bar{x},\bar{y}) = 0\}.$$

The need to assume that $(\bar{x},\bar{y},\bar{u})$ is a local optimal solution for all $\bar{u} \in \Lambda(\bar{x},\bar{y})$ is forced by the complementarity slackness conditions which reduce **gph** Ψ to a proper subset if one component of the Lagrange multiplier $\bar{u}_i$ is restricted to take on a positive value, see Example 1.2 below.

Problem (3) is a so-called *mathematical program with complementarity constraints* (MPCC) (or mathematical program with equilibrium constraints (MPEC)), see the monographs of Luo et al. [14] or Outrata et al. [15]. Thus, necessary optimality conditions for the (MPCC) can also be used as necessary optimality conditions for the bilevel optimization problem. Note that MFCQ is violated at every feasible point of (3) as Scheel and Scholtes have shown in [16].

To overcome the resulting numerical difficulties that arise when solving the (MPCC), Scholtes [17] suggested to solve a sequence of optimization problems

$$\begin{aligned} F(x,y) &\to \min_{x,y,u} \\ G(x) &\le 0 \\ \nabla_y L(x,y,u) &= 0 \\ g(x,y) &\le 0 \\ u &\ge 0 \\ -u_i g_i(x,y) &\le \varepsilon,\ i = 1,\dots,p \end{aligned} \tag{4}$$

for $\varepsilon \searrow 0$. He has shown that a sequence of stationary points $\{(x_\varepsilon, y_\varepsilon, u_\varepsilon)\}$ of problem (4) converges to a so-called C-stationary point $(\bar{x}, \bar{y}, \bar{u})$ of problem (3) for $\varepsilon \searrow 0$ provided that the so-called MPEC-LICQ is satisfied at $(\bar{x}, \bar{y}, \bar{u})$, where LICQ is an abbreviation of the linear independence constraint qualification and MPEC-LICQ is a version of the LICQ especially adapted to the (MPCC). The definition of C-stationarity can be found in Section 4. It has been shown by Hoheisel et al. [18] that this result is correct even under the weaker MPEC-MFCQ. Unfortunately, local optimal solutions of problem (3) are generally not related to local optimal points of the bilevel optimization problem, as the following example shows.

Example 1.2 (Dempe and Dutta [13]): Consider the bilevel optimization problem

$$\min\{(x-1)^2 + (y-1)^2 : (x,y) \in \mathbf{gph}\ \Psi\}, \tag{5}$$

where

$$\Psi(x) = \operatorname*{Argmin}_y \{-y : x + y \le 1,\ -x + y \le 1\}. \tag{6}$$

The unique optimal solution $y(x)$ of the lower level problem and the set $\Lambda(x,y)$ of Lagrange multipliers are given by

$$y(x) = \begin{cases} x+1 & \text{if } x \le 0, \\ -x+1 & \text{if } x \ge 0, \end{cases} \quad \Lambda(x,y) = \begin{cases} \{(1,0)\} & \text{if } x > 0, \\ \{(0,1)\} & \text{if } x < 0, \\ \mathbf{conv}\,\{(1,0),(0,1)\} & \text{if } x = 0, \end{cases}$$

where **conv** A denotes the convex hull of the set A. The bilevel optimization problem has the unique optimal solution $(\bar{x}, \bar{y}) = (0.5, 0.5)$ and no other local optimal solutions. The points $(\bar{x}, \bar{y}, \bar{u}_1, \bar{u}_2) = (0.5, 0.5, 1, 0)$ and $(\hat{x}, \hat{y}, \hat{u}_1, \hat{u}_2) = (0, 1, 0, 1)$ are local optimal solutions of problem (3). The point $(0, 1, 0, 1)$ is not related to a local optimal solution of (5).

Mersha [19] considered another numerically stable approach, namely a sequence of perturbed problems

$$\begin{aligned} F(x,y) &\to \min_{x,y,u} \\ G(x) &\leq 0 \\ \|\nabla_y L(x,y,u)\| &\leq \varepsilon_1 \\ g(x,y) &\leq 0 \\ u &\geq 0 \\ -u_i g_i(x,y) &\leq \varepsilon_2,\ i = 1,\dots,p \end{aligned} \tag{7}$$

for $(\varepsilon_1, \varepsilon_2) \searrow (0,0)$ (cf. also the article [20] of Mersha and Dempe). He has shown that a sequence of local optimal points $\{(x^k, y^k, u^k)\}_{k=1}^{\infty}$ of (7) for $(\varepsilon_1^k, \varepsilon_2^k) \searrow (0,0)$ converges to a B-stationary point $(\bar{x}, \bar{y}, \bar{u})$ of (1), (2) provided that the MFCQ, the constant rank constraint qualification and the strong sufficient optimality condition of second order are satisfied for the lower level problem (1) at $\bar{y}$ for $x = \bar{x}$. Here, a point $(\bar{x}, \bar{y}, \bar{u})$ is B-stationary if there does not exist a feasible descent direction for problem (2) at $\bar{x}$ if y is replaced by the function $y \mapsto y(x)$ in this problem. An essential part of the Mersha's proof in [19] is that the optimal solution $y = y(x)$ of the lower level problem (1) is strongly stable in the sense of Kojima [21] and directionally differentiable under these assumptions, see Ralph and Dempe [22].

The assumptions for strong stability of the solution of the lower level problem are rather restrictive. Moreover, the algorithm proposed in [19, 20] implements a method of feasible directions. To compute a descent direction, a bilevel optimization problem has to be solved where the lower level problem is the quadratic optimization problem for computing the directional derivative of the function $y(x)$ [22]. An active set strategy is used to replace the complementarity slackness conditions of the KKT conditions of the lower level problem which realizes some combinatorial ideas to compute the desired direction. This method is, thus, rather expensive. We will show in this article that the assumptions used in [19, 20] are much too restrictive and can substantially be weakened maintaining the result that locally optimal solutions of the bilevel optimization problem can be computed solving problem (7). Moreover, we do not need to solve this problem applying an especially tailored algorithm but can use an arbitrary one.

In Section 2 it is shown that a sequence of global optima of problem (7) converges to a global optimum of the bilevel optimization problem. Local optima will be investigated in Section 3. Topic of Section 4 is the proof that a sequence of local optima of (7) converges to a C-stationary point of problem (3). The use of variational analysis makes this proof elegant and clear.

2. Global optimal solutions of problem (7)

Consider problem (7) with Manhattan norm $\|a\|_1 = \sum_{i=1}^n |a_i|$, Euclidean norm $\|a\|_2 = \sqrt{\sum_{i=1}^n a_i^2}$ or Chebyshev norm $\|a\|_\infty = \max_{i=1,\dots,n} |a_i|$. Then it is clear that every feasible point (x, y, u) of problem (3) is feasible for (7) with $\varepsilon_1 \geq 0$, $\varepsilon_2 \geq 0$, too. Hence, the global optimal function value of problem (7) is not larger than the one of problem (3). Moreover, if $\{(x^k, y^k, u^k)\}_{k=1}^\infty$ is a sequence of global optima of problem (7) for $(\varepsilon_1^k, \varepsilon_2^k) \searrow (0, 0)$ and this sequence converges to $(\bar{x}, \bar{y}, \bar{u})$, the limit point $(\bar{x}, \bar{y}, \bar{u})$ is feasible for (3). This implies that this point is a global optimal solution of (3). Using the result of [13] that global optimal solutions of (3) correspond to global optimal solutions of (1), (2), the following theorem has been shown:

Theorem 2.1. *Consider problem (7) for $\{(\varepsilon_1^k, \varepsilon_2^k)\}_{k=1}^\infty \subset \mathbb{R}_+^2$ tending to $(0, 0)$, and let $\{(x^k, y^k, u^k)\}_{k=1}^\infty$ be a sequence of global optimal solutions of this problem for $(\varepsilon_1, \varepsilon_2) = (\varepsilon_1^k, \varepsilon_2^k)$. Then, the part $(\bar{x}, \bar{y})$ of any accumulation point $(\bar{x}, \bar{y}, \bar{u})$ of $\{(x^k, y^k, u^k)\}_{k=1}^\infty$ is a global optimal solution of (1), (2).*

Note that the problems (7) have global optimal solutions if the set

$$M := \{(x, y) : g(x, y) \leq 0,\ G(x) \leq 0\}$$

is nonempty and compact and MFCQ is satisfied for the lower level problem at all points $(x, y) \in M$. This is also an assumption guaranteeing the existence of an optimal solution of the optimistic bilevel optimization problem (1), (2), see e.g. Lignola and Morgan [23] or Dempe [2].

3. Local optimal solutions of problem (7)

3.1. A direct approach

In this section consider problem (7) with Chebyshev norm (for simplicity) and assume that MFCQ is satisfied for the lower level problem at all feasible points. Note that this allows us to equivalently rewrite the constraint $\|\nabla_y L(x, y, u)\|_\infty \leq \varepsilon_1$ as $-\varepsilon_1 \leq \nabla_y L(x, y, u) \leq \varepsilon_1$, meaning that the value of every component of the vector $\nabla_y L(x, y, u)$ belongs to $[-\varepsilon_1, \varepsilon_1]$. This will significantly simplify our further notations since we do not need to use subdifferential calculus.

Definition 3.1: Consider a local optimal solution $(\bar{x}, \bar{y}, \bar{u})$ of (7) with sufficiently small $\varepsilon_i > 0$, $i = 1{,}2$. The point is a strong local solution of this problem if there exists $\gamma > 0$ such that $F(x, y) \geq F(\bar{x}, \bar{y})$ for all feasible points (x, y, u) of (7) with sufficiently small $\varepsilon_i > 0$, $i = 1{,}2$ and $\|(x, y, u) - (\bar{x}, \bar{y}, \bar{u})\|_2 \leq \gamma$.

Note that the parameter γ is independent of $\varepsilon_1, \varepsilon_2$.

Theorem 3.2. *Consider problem (1), (2), and let MFCQ be satisfied at all $(x, y) \in M$. Let $(\bar{x}, \bar{y}, \bar{u})$ be an accumulation point of a sequence $\{(\bar{x}^t, \bar{y}^t, \bar{u}^t)\}_{t=1}^{\infty}$ of local optimal solutions of the problems (7) for $\{(\varepsilon_1^t, \varepsilon_2^t)\}_{t=1}^{\infty} \subset \mathbb{R}_+^2$ tending to $(0, 0)$. Assume that all $\{(\bar{x}^t, \bar{y}^t, \bar{u}^t)\}_{t=1}^{\infty}$ are strong local optimal solutions with the same $\gamma > 0$. Then, $(\bar{x}, \bar{y})$ is a local optimal solution of problem (1), (2).*

Proof: Let $(\bar{x}, \bar{y}, \bar{u})$ be an accumulation point of a sequence $\{(\bar{x}^t, \bar{y}^t, \bar{u}^t)\}_{t=1}^{\infty}$ of local optimal solutions of the problems (7) for $\{(\varepsilon_1^t, \varepsilon_2^t)\}_{t=1}^{\infty} \subset \mathbb{R}_+^2$ tending to $(0, 0)$, and assume that $(\bar{x}, \bar{y})$ is not a local optimum of (1), (2). Then, due to Theorem 1.1, there exists $\tilde{u} \in \Lambda(\bar{x}, \bar{y})$ such that $(\bar{x}, \bar{y}, \tilde{u})$ is not a local optimum of (3). Thus, there exists a sequence $\{(x^k, y^k, u^k)\}_{k=1}^{\infty}$ of feasible points of (3) converging to $(\bar{x}, \bar{y}, \tilde{u})$ with $F(x^k, y^k) < F(\bar{x}, \bar{y})$ for all k.

By the properties of the sequence $\{(x^k, y^k, u^k)\}_{k=1}^{\infty}$, we deduce that $g(x^k, y^k) \leq 0$ and $G(x^k) \leq 0$ for all k.

Consider the sequence $\{(\bar{x}^t, \bar{y}^t, \tilde{u})\}_{t=1}^{\infty}$. This sequence converges to $(\bar{x}, \bar{y}, \tilde{u})$. Hence, for each $(\varepsilon_1^t, \varepsilon_2^t) > 0$, there exists a $t_0(\varepsilon_1^t, \varepsilon_2^t)$ such that $(\bar{x}^t, \bar{y}^t, \tilde{u})$ is feasible for problem (7) for $(\varepsilon_1^t, \varepsilon_2^t)$ for all $t \geq t_0(\varepsilon_1^t, \varepsilon_2^t)$. Let $t = t_0(\varepsilon_1^t, \varepsilon_2^t)$ and renumber the sequence $\{(\bar{x}^t, \bar{y}^t, \tilde{u})\}_{t=1}^{\infty}$ in that sense without loss of generality.

Then, since the objective function of problem (7) does not depend on $\tilde{u}$, $(\bar{x}^t, \bar{y}^t, \tilde{u})$ is a local optimal solution of this problem and, by the assumptions of the theorem, a strong local optimum.

By convergence of $\{(\bar{x}^t, \bar{y}^t, \tilde{u})\}_{t=1}^{\infty}$ to $(\bar{x}, \bar{y}, \tilde{u})$ there is t_0 such that $\|(\bar{x}^t, \bar{y}^t, \tilde{u}) - (\bar{x}, \bar{y}, \tilde{u})\|_2 \leq \gamma/2$ for all $t \geq t_0$. Also, by $\{(x^k, y^k, u^k)\}_{k=1}^{\infty}$ converging to $(\bar{x}, \bar{y}, \tilde{u})$, there is k_t such that $\|(x^{k_t}, y^{k_t}, u^{k_t}) - (\bar{x}, \bar{y}, \tilde{u})\|_2 \leq \gamma/2$ and $(x^{k_t}, y^{k_t}, u^{k_t})$ is feasible for (7) with $(\varepsilon_1, \varepsilon_2) = (\varepsilon_1^t, \varepsilon_2^t)$. Hence,

$$\|(x^{k_t}, y^{k_t}, u^{k_t}) - (\bar{x}^t, \bar{y}^t, \tilde{u})\|_2 \leq \gamma$$

implying that

$$F(\bar{x}^t, \bar{y}^t) \leq F(x^{k_t}, y^{k_t}) < F(\bar{x}, \bar{y}) \; \forall t \tag{8}$$

since $(\bar{x}^t, \bar{y}^t, \tilde{u})$ is a strong local optimum.

Next, we show that this implies that $(\bar{x}, \bar{y}, \tilde{u})$ is a local optimum of (3), which violates the assumption.

Let $(x^0, y^0) \in \mathbf{gph}\, \Psi$, $G(x^0) \leq 0$, $u^0 \in \Lambda(x^0, y^0)$ be an arbitrarily chosen feasible point for problem (3) with $\|(x^0, y^0, u^0) - (\bar{x}, \bar{y}, \tilde{u})\|_2 \leq \gamma/2$. (x^0, y^0, u^0) is feasible for problem (7) for $(\varepsilon_1, \varepsilon_2) \geq 0$. Since $(\bar{x}^t, \bar{y}^t, \tilde{u})$ converges to $(\bar{x}, \bar{y}, \tilde{u})$, we have $\|(\bar{x}^t, \bar{y}^t, \tilde{u}) - (\bar{x}, \bar{y}, \tilde{u})\|_2 \leq \gamma/2$ and, thus, $\|(\bar{x}^t, \bar{y}^t, \tilde{u}) - (x^0, y^0, u^0)\|_2 \leq \gamma$ for sufficiently large t. This implies

$$F(\bar{x}^t, \bar{y}^t) \leq F(x^0, y^0)$$

for all t sufficiently large by the assumptions of the theorem. Passing to the limit for $t \to \infty$, $(\bar{x}, \bar{y}, \tilde{u})$,

$$F(\bar{x}, \bar{y}) \leq F(x^0, y^0)$$

is shown. Thus, $(\bar{x}, \bar{y})$ is a local minimum of (3), contradicting our initial assumption. Hence, $\tilde{u}$ such that $(\bar{x}, \bar{y}, \tilde{u})$ is not a local minimum of (3) does not exist and $(\bar{x}, \bar{y})$ is a local minimum of (1), (2) by the results of Dempe and Dutta in [13]. ■

Remark 3.3: The idea to use neighbourhoods of feasible points of the (MPCC) (3) excluding the Lagrange multiplier of the lower level problem can also be found in [24, 25].

3.2. Strongly stable optimal solutions in the lower level problem

The following result can be found in [26]:

Theorem 3.4 (Mersha and Dempe [26])**.** *Consider problem* (1), (2). *Regarding the lower level, let besides MFCQ also the constant rank constraint qualification (CRCQ):*

there is an open neighbourhood $U(\bar{x}, \bar{y})$ of the point $(\bar{x}, \bar{y})$ such that for each subset $I \subseteq \{i : g_i(\bar{x}, \bar{y}) = 0\}$, the family of gradient vectors $\{\nabla_y g_i(x, y) : i \in I\}$ has the same rank for all $(x, y) \in U(\bar{x}, \bar{y})$

and the strong sufficient optimality condition of second order (*SSOSC*):

for each $u \in \Lambda(\bar{x}, \bar{y})$ and for every nonzero element of the set

$$\left\{r \in \mathbb{R}^m : \nabla_y g_i(\bar{x}, \bar{y}) r = 0,\ i \in \{j : u_j > 0\}\right\},$$

we have

$$r^\top \nabla^2_{yy} L(\bar{x}, \bar{y}, u) r > 0$$

*be satisfied at all points $(\bar{x}, \bar{y}) \in$ **gph** Ψ. Let $\{(x^k, y^k, u^k)\}_{k=1}^\infty$ be a sequence of local optimal solutions of problem* (7) *for $\{(\varepsilon_1^k, \varepsilon_2^k)\}_{k=1}^\infty \subset \mathbb{R}^2_+$ converging to $(0, 0)$. Then, any accumulation point (x^*, y^*, u^*) of the sequence $\{(x^k, y^k, u^k)\}_{k=1}^\infty$ is a Bouligand stationary point of problem* (3), *i.e. there does not exist a feasible descent direction provided that*

$$\left\{(d, r) \in \mathbb{R}^n \times \mathbb{R}^m : \begin{array}{l} \nabla G_i(x^*) d < 0,\ i : G_i(x^*) = 0, \\ \nabla g_j(x^*, y^*)(d, r)^\top < 0,\ j \in \{j : g_j(x^*, y^*) = 0\} \end{array}\right\} \neq \emptyset.$$

The optimal solution $y(x)$ of the parametric optimization problem (1) is a piecewise continuously differentiable vector-valued function (a so-called PC^1 function), and it is directionally differentiable if the assumptions of Theorem 3.4 are satisfied (cf. Ralph and Dempe [22]). To compute the directional derivative $y'(x; d)$, a quadratic optimization problem needs to be solved (cf. Ralph and Dempe [22]) which has a unique optimal solution. The KKT conditions of this problem are part of the constraints of the following problem which is a linear approximation of (7) with $\varepsilon_i = 0,\ i = 1, 2$:

$$\begin{aligned} \nabla F(x,y)(d,r)^\top &\to \min_{d,r,v} \\ \nabla G_i(x)d &\le 0,\ i \in K(x) := \{i : G_i(x) = 0\} \\ \|\nabla(\nabla_y L(x,y,u))(d,r,v)^\top\| &\le \varepsilon_1 \\ \nabla g_i(x,y)(d,r)^\top &\le 0,\ i \in I(x,y) := \{i : g_i(x,y) = 0\} \\ v_i &\ge 0,\ i \in J(u) := \{i : u_i = 0\} \\ -\nabla(u_i g_i(x,y))(d,r,v)^\top &\le \varepsilon_2,\ i \in \{i : u_i g_i(x,y) = \varepsilon_2\} \\ \|(d,r)\| &= 1. \end{aligned} \tag{9}$$

The point $(\bar{x}, \bar{y}, \bar{u})$ is a Bouligand stationary point of (7) for $\varepsilon_i = 0,\ i = 1, 2$ if the optimal function value of problem (9) with $\varepsilon_i = 0,\ i = 1, 2$ is zero.

Using Theorem 3.4, the following can easily be shown.

Theorem 3.5. *Let the assumptions of Theorem 3.4 be satisfied, let the sequence $\{(\varepsilon_1^k, \varepsilon_2^k)\}_{k=1}^\infty \subset \mathbb{R}^2_+$ converge to $(0,0)$ and let $(\bar{x}, \bar{y}, \bar{u})$ be an accumulation point of a sequence $\{(x^k, y^k, u^k)\}_{k=1}^\infty$ of local optimal solutions of the problems (7) for $\{(\varepsilon_1^k, \varepsilon_2^k)\}_{k=1}^\infty$. Then, $(\bar{x}, \bar{y})$ is a Bouligand stationary point of (1), (2), i.e.*

$$\nabla_x F(\bar{x}, \bar{y})d + \nabla_y F(\bar{x}, \bar{y})y'(\bar{x}; d) \ge 0$$

for all d with $\nabla G_i(\bar{x})d \le 0,\ i \in K(\bar{x})$.

Corollary 3.6. *Let the assumptions of Theorem 3.4 be satisfied and assume that*

$$\nabla_x F(\bar{x}, \bar{y})d + \nabla_y F(\bar{x}, \bar{y})y'(\bar{x}; d) > 0$$

for all d with $\nabla G_i(\bar{x})d \le 0,\ i \in K(\bar{x})$ for all Bouligand stationary points. Then, the accumulation points $(\bar{x}, \bar{y}, \bar{u})$ of a sequence $\{(x^k, y^k, u^k)\}_{k=1}^\infty$ of local optimal solutions of the problems (7) for $\{(\varepsilon_1^k, \varepsilon_2^k)\}_{k=1}^\infty \subset \mathbb{R}^2_+$ tending to $(0,0)$ correspond to local optimal solutions $(\bar{x}, \bar{y})$ of problem (1), (2).

3.3. Locally unique local optimal solution

We will need the following definitions for our further considerations.

Definition 3.7: Consider a point-to-set mapping $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ and $\bar{x} \in \mathbb{R}^n$. A is called locally compact at $\bar{x}$ if for a compact set $\Omega \supset A(\bar{x})$, there is a neighbourhood $U(\bar{x})$ of $\bar{x}$ with $A(x) \subset \Omega$ for all $x \in U(\bar{x})$.

Definition 3.8: A point-to-set mapping $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is lower semicontinuous (in the sense of Berge) at $\bar{x} \in \mathbb{R}^n$ if for every open set Ω with $\Omega \cap A(\bar{x}) \ne \emptyset$, there exists a neighbourhood $U(\bar{x})$ of $\bar{x}$ such that $A(x) \cap \Omega \ne \emptyset$ for all $x \in U(\bar{x})$.

The point-to-set mapping A is upper semicontinuous (in the sense of Berge) at $\bar{x}$ if for every open set $\Omega \supseteq A(\bar{x})$, there exists a neighbourhood $U(\bar{x})$ of $\bar{x}$ such that $A(x) \subset \Omega$ for all $x \in U(\bar{x})$.

Let

$$P(\varepsilon_1, \varepsilon_2) := \{(x, y, u) : \; G(x) \leq 0, \; g(x, y) \leq 0, \; u \geq 0, \\ \|\nabla_y L(x, y, u)\|_\infty \leq \varepsilon_1, -u_i g_i(x, y) \leq \varepsilon_2, \; i = 1, \dots, p\}$$

be the feasible set of problem (7). Let MFCQ be satisfied at all $(x, y) \in M$. Then,

$$P(0, 0) = \{(x, y, u) : G(x) \leq 0, \; (x, y) \in \mathbf{gph}\, \Psi, \; u \in \Lambda(x, y)\}$$

and $\Lambda(x, y)$ are compact.

Theorem 3.9. *Let MFCQ be satisfied at all $(x, y) \in M$ which is assumed to be a nonempty and compact set. Then, $P : \mathbb{R}^2_+ \rightrightarrows \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$ is continuous at $(0, 0)$ with respect to $\mathbb{R}^2_+$.*

Proof: We have

$$P(\varepsilon_1, \varepsilon_2) \subseteq M \times \mathbb{R}^p.$$

Furthermore, $\Lambda(\bar{x}, \bar{y})$ is nonempty and compact for all $(\bar{x}, \bar{y}) \in \mathbf{gph}\, \Psi$ by [27] (due to the assumed constraint qualification), which also gives us the local compactness of $\Lambda(\cdot, \cdot)$. This allows us to apply [28, Theorem 3.1.2(i)] in order to deduce that the mapping $(x, y) \rightrightarrows \Lambda(x, y)$ is upper semicontinuous at $(\bar{x}, \bar{y})$. Since $\mathbf{gph}\, \Psi \subset M$ and M is compact, we deduce that $P(0, 0)$ is bounded. Using again [28, Theorem 3.1.2], we can conclude that $P : \mathbb{R}^2_+ \rightrightarrows \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$ is upper semicontinuous at $(0, 0)$. Since the mapping $P : \mathbb{R}^2_+ \rightrightarrows \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$ is increasing in the sense that

$$P(\varepsilon_1, \varepsilon_2) \subseteq P(\bar{\varepsilon}_1, \bar{\varepsilon}_2) \; \forall \, \bar{\varepsilon}_i \geq \varepsilon_i \geq 0, \; i = 1, 2$$

and, hence, $P(0, 0) \subseteq P(\varepsilon_1, \varepsilon_2)$ for all $(\varepsilon_1, \varepsilon_2) \in \mathbb{R}^2_+$, it is also lower semicontinuous. ■

Using parametric optimization once more, we can exploit the stability of the feasible set mapping of problem (7) to prove that local optimal solutions can be computed using a sequence of problems (7) for decreasing $(\varepsilon_1, \varepsilon_2) \in \mathbb{R}^2_+$ provided that a local optimal solution of problem (1), (2) is locally unique.

Definition 3.10: Let $(\bar{x}, \bar{y}) \in M$ be a local optimal solution of problem (1), (2). The point is locally unique if there is a sufficiently small open neighbourhood $U(\bar{x}, \bar{y})$ of $(\bar{x}, \bar{y})$ such that $F(x, y) > F(\bar{x}, \bar{y})$ for all $(x, y) \in \mathbf{gph}\, \Psi \cap U(\bar{x}, \bar{y})$, $(x, y) \neq (\bar{x}, \bar{x})$ with $G(x) \leq 0$.

Theorem 3.11. *Let the assumptions of Theorem 3.9 be satisfied, let $(\bar{x}, \bar{y})$ be a local optimal solution of problem (1), (2) and assume that $(\bar{x}, \bar{y})$ is locally unique. Then, the function*

$$v(\varepsilon_1, \varepsilon_2) := \min_{x,y,u} \{F(x, y) : (x, y, u) \in P(\varepsilon_1, \varepsilon_2) \cap \left(U(\bar{x}, \bar{y}) \times \mathbb{R}^p\right)\}$$

is continuous at $(0, 0)$ with respect to $\mathbb{R}^2_+$, and the mapping

$$S(\varepsilon_1, \varepsilon_2) := \{(x, y) : \exists u \text{ such that } (x, y, u) \in P(\varepsilon_1, \varepsilon_2) \cap \left(U(\bar{x}, \bar{y}) \times \mathbb{R}^p\right) : \\ F(x, y) \leq v(\varepsilon_1, \varepsilon_2)\}$$

is continuous at $(0, 0)$ with respect to $\mathbb{R}^2_+$. Here, $U(\bar{x}, \bar{y})$ is the open neighbourhood from Definition 3.10.

Proof: Clearly, the mapping $(\varepsilon_1, \varepsilon_2) \rightrightarrows P(\varepsilon_1, \varepsilon_2) \cap (U(\bar{x}, \bar{y}) \times \mathbb{R}^p)$ is continuous due to Theorem 3.9. According to the assumptions of the theorem, $S(0, 0) = \{(\bar{x}, \bar{y})\}$ is a compact set. Then, the continuity of the function v at $(0, 0)$ with respect to $\mathbb{R}^2_+$ follows from [28, Theorem 4.2.2]. The same is true for the upper semicontinuity of the mapping S at $(0, 0)$.

Let $\bar{u} \in \Lambda(\bar{x}, \bar{y})$ and $\{(\varepsilon_1^k, \varepsilon_2^k)\}_{k=1}^\infty \subset \mathbb{R}^2_+$ be an arbitrary sequence converging to $(0, 0)$. By the lower semicontinuity of the mapping P at $(0, 0)$, there exists a sequence $\{(x^k, y^k, u^k)\}_{k=1}^\infty$ converging to $(\bar{x}, \bar{y}, \bar{u})$ with $(x^k, y^k, u^k) \in P(\varepsilon_1^k, \varepsilon_2^k)$ for all k. Moreover, there exists a sequence $\{(\tilde{x}^k, \tilde{y}^k, \tilde{u}^k)\}_{k=1}^\infty$ with $(\tilde{x}^k, \tilde{y}^k, \tilde{u}^k) \in P(\varepsilon_1^k, \varepsilon_2^k)$, $(\tilde{x}^k, \tilde{y}^k) \in S(\varepsilon_1^k, \varepsilon_2^k)$. Then, we have

$$v(0, 0) = \lim_{k\to\infty} v(\varepsilon_1^k, \varepsilon_2^k) = \lim_{k\to\infty} F(\tilde{x}^k, \tilde{y}^k) \leq \lim_{k\to\infty} F(x^k, y^k) = F(\bar{x}, \bar{y}) = v(0, 0).$$

Hence, the accumulation points of the sequence $\{(\tilde{x}^k, \tilde{y}^k)\}_{k=1}^\infty$ are elements of $S(0, 0)$, and since $S(0, 0)$ reduces to a singleton, $\lim_{k\to\infty}(\tilde{x}^k, \tilde{y}^k) = (\bar{x}, \bar{y})$. This implies the lower semicontinuity of the mapping S. ■

The local uniqueness of the local optimal solution is obviously a sufficient condition for strong local optimality from Section 3.1.

Proposition 3.12. *Let $\varepsilon_i > 0$, $i = 1, 2$. Assume that there is a neighbourhood $U(\bar{x}, \bar{y})$ of $(\bar{x}, \bar{y})$ such that $F(\bar{x}, \bar{y}) < F(x, y)$ for all $(x, y, u) \in P(\varepsilon_1, \varepsilon_2) \cap U(\bar{x}, \bar{y}) \times \mathbb{R}^p$, $(x, y) \neq (\bar{x}, \bar{y})$, where $u \in \mathbb{R}^p$ is arbitrary chosen such that $(x, y, u) \in P(\varepsilon_1, \varepsilon_2)$. Then, $(\bar{x}, \bar{y}, u)$ is a strong local optimal solution of (3).*

Proposition 3.13. *Let $\Gamma(\bar{x}, \bar{y}) \subseteq \mathbb{R}^{n+m}$ be a cone containing the Bouligand (or tangent) cone at $(\bar{x}, \bar{y})$ with respect to the set* **gph** *$\Psi \cap \{(x, y) : G(x) \leq 0\}$. If $\nabla F(\bar{x}, \bar{y})(d, r)^\top > 0$ for all $(d, r) \in \Gamma(\bar{x}, \bar{y})$, $(d, r) \neq (0, 0)$, then $(\bar{x}, \bar{y}, \bar{u})$ is a strong local optimal solution of (7) for arbitrary $\bar{u} \in \Lambda(\bar{x}, \bar{y})$.*

Proof: Assume that $(\bar{x}, \bar{y}, \bar{u})$ is not a strong local optimal solution of (7) for arbitrary $\bar{u} \in \Lambda(\bar{x}, \bar{y})$. Then, for each $\gamma_k > 0$, there exists (x^k, y^k, u^k), $u^k \in \Lambda(x^k, y^k)$ with $\|(x^k, y^k, u^k) - (\bar{x}, \bar{y}, \bar{u})\|_2 \leq \gamma_k$ and $F(x^k, y^k) < F(\bar{x}, \bar{y})$. By upper semicontinuity, every accumulation point of $\{u^k\}_{k=1}^{\infty}$ belongs to $\Lambda(\bar{x}, \bar{y})$. Moreover, $\{(x^k, y^k) - (\bar{x}, \bar{y})/\|(x^k, y^k) - (\bar{x}, \bar{y})\|_2\}_{k=1}^{\infty}$ is bounded and has nonzero accumulation points $(d, r) \in \Gamma(\bar{x}, \bar{y})$. Due to $F(x^k, y^k) < F(\bar{x}, \bar{y})$ for all k, we have $\nabla F(\bar{x}, \bar{y})(d, r)^\top \leq 0$, contradicting the assumption. ■

4. C-stationary points

Next, we investigate accumulation points of sequences of stationary points of the problems (7). Convergence of an algorithm using problem (4) to a C-stationary solution of (3) can be found in Hoheisel et al. [18]. Using the same ideas as in [18] convergence of an algorithm using (7) to a C-stationary point of (3) can be verified. In the following we give an alternative proof of this result using variational analysis. In our opinion this proof is much more clear and elegant. Using the Chebyshev norm, problem (7) can be reformulated as

$$
\begin{aligned}
F(x,y) &\to \min_{x,y,u} \\
G(x) &\leq 0 \\
-\varepsilon_1 \leq \nabla_{y_i} L(x,y,u) &\leq \varepsilon_1, \; i = 1, \ldots, m \\
g(x,y) &\leq 0 \\
u &\geq 0 \\
-u_i g_i(x,y) &\leq \varepsilon_2, \; i = 1, \ldots, p.
\end{aligned}
\tag{10}
$$

Let (x, y, u) be a feasible point of this problem, and consider the index sets $K(x)$, $I(x, y)$ and $J(u)$ from (9) as well as

$$
\begin{aligned}
I_-(x,y,u,\varepsilon_1) &:= \{i \in \{1,\ldots,m\} : \nabla_{y_i} L(x,y,u) = -\varepsilon_1\}, \\
I_+(x,y,u,\varepsilon_1) &:= \{i \in \{1,\ldots,m\} : \nabla_{y_i} L(x,y,u) = \varepsilon_1\}, \\
I_{gu}(x,y,u,\varepsilon_2) &:= \{i \in \{1,\ldots,p\} : -u_i g_i(x,y) = \varepsilon_2\}.
\end{aligned}
$$

For a feasible point (x, y, u) of (3), we furthermore define the index sets

$$
\begin{aligned}
I_g(x,y,u) &:= \{i \in \{1,\ldots,p\} : \; g_i(x,y) = 0, u_i > 0\}, \\
I_u(x,y,u) &:= \{i \in \{1,\ldots,p\} : \; g_i(x,y) < 0, u_i = 0\}, \\
I_{gu}(x,y,u) &:= \{i \in \{1,\ldots,p\} : \; g_i(x,y) = 0, u_i = 0\}.
\end{aligned}
$$

A feasible point $(\bar{x}, \bar{y}, \bar{u})$ of problem (3) satisfies MPEC-MFCQ if there does not exist $(\alpha, \beta, \gamma, \delta) \neq (0, 0, 0, 0)$ such that

$$\begin{pmatrix} \nabla G(\bar{x})^\top \alpha \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \nabla_x g(\bar{x}, \bar{y})^\top \beta \\ \nabla_y g(\bar{x}, \bar{y})^\top \beta \\ 0 \end{pmatrix} + \begin{pmatrix} \nabla_x (\nabla_y L(\bar{x}, \bar{y}, \bar{u}))^\top \gamma \\ \nabla_y (\nabla_y L(\bar{x}, \bar{y}, \bar{u}))^\top \gamma \\ \nabla_y g(\bar{x}, \bar{y}) \gamma \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ -\delta \end{pmatrix} = 0,$$

$$\alpha \geq 0,\ \alpha^\top G(\bar{x}) = 0,\ \beta_i = 0,\ i \in I_u(\bar{x}, \bar{y}, \bar{u}),\ \delta_i = 0,\ i \in I_g(\bar{x}, \bar{y}, \bar{u}).$$

Definition 4.1: A feasible point $(\bar{x}, \bar{y}, \bar{u})$ of (3) is a C-stationary point of this problem if there exists $(\alpha, \beta, \gamma, \delta) \in \mathbb{R}^{q+p+m+p}$ such that

$$\begin{pmatrix} \nabla_x F(\bar{x}, \bar{y})^\top \\ \nabla_y F(\bar{x}, \bar{y})^\top \\ 0 \end{pmatrix} + \begin{pmatrix} \nabla G(\bar{x})^\top \alpha \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \nabla_x g(\bar{x}, \bar{y})^\top \beta \\ \nabla_y g(\bar{x}, \bar{y})^\top \beta \\ 0 \end{pmatrix}$$
$$+ \begin{pmatrix} \nabla_x (\nabla_y L(\bar{x}, \bar{y}, \bar{u}))^\top \gamma \\ \nabla_y (\nabla_y L(\bar{x}, \bar{y}, \bar{u}))^\top \gamma \\ \nabla_y g(\bar{x}, \bar{y}) \gamma \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ -\delta \end{pmatrix} = 0 \tag{11}$$

with

$$\alpha \geq 0,\ \alpha^\top G(\bar{x}) = 0, \tag{12}$$

$$\beta^\top g(\bar{x}, \bar{y}) = 0, \tag{13}$$

$$\beta_i = 0,\ i \in I_u(\bar{x}, \bar{y}, \bar{u}), \tag{14}$$

$$\delta_i = 0,\ i \in I_g(\bar{x}, \bar{y}, \bar{u}), \tag{15}$$

$$\beta_i \delta_i \geq 0,\ i \in I_{gu}(\bar{x}, \bar{y}, \bar{u}). \tag{16}$$

Theorem 4.2. *Assume that $\{(x^k, y^k, u^k)\}_{k=1}^\infty$ is a sequence of local optimal solutions of (10) for parameters $\{(\varepsilon_1^k, \varepsilon_2^k)\}_{k=1}^\infty \subset \mathbb{R}_+^2$ converging to $(0, 0)$. Let the accumulation point of $\{(x^k, y^k, u^k)\}_{k=1}^\infty$ be $(\bar{x}, \bar{y}, \bar{u})$. Furthermore, suppose that MPEC-MFCQ holds at $(\bar{x}, \bar{y}, \bar{u})$ for (3). Then, $(\bar{x}, \bar{y}, \bar{u})$ is a C-stationary point of (3).*

Proof: The point (x^k, y^k, u^k) being an optimal solution of (10) for some $(\varepsilon_1, \varepsilon_2) = (\varepsilon_1^k, \varepsilon_2^k)$ is equivalent to $(x^k, y^k, u^k, a^k, b^k)$ with $a^k = -g(x^k, y^k)$, $b^k = u^k$ solving

$$\begin{array}{rcll} F(x, y) & \rightarrow & \min_{x,y,u,a,b} & \\ G(x) & \leq & 0 & \\ -\varepsilon_1 \leq \nabla_{y_i} L(x, y, u) & \leq & \varepsilon_1, & i = 1, \ldots, m \\ g(x, y) + a & = & 0 & \\ -u + b & = & 0 & \\ (a, b) & \in & \Omega_{\varepsilon_2} := & \big\{(\omega_1, \omega_2) \in \mathbb{R}^p \times \mathbb{R}^p : \\ & & & \omega_1 \geq 0, \omega_2 \geq 0, \omega_{1,i}\omega_{2,i} \\ & & & \quad \leq \varepsilon_2,\ i = 1, \ldots, p\big\}. \end{array}$$

This means that there exist $(\chi^k,\alpha^k,\beta^k,\gamma_1^k,\gamma_2^k,\delta^k)\neq(0,0,0,0,0)$, $\chi^k\geq 0$ and $(a^{*k},b^{*k})\in N_{\Omega_{\varepsilon_2^k}}(a^k,b^k)$ such that

$$\chi^k\begin{pmatrix}\nabla_x F(x^k,y^k)^\top\\ \nabla_y F(x^k,y^k)^\top\\ 0\\ 0\\ 0\end{pmatrix}+\begin{pmatrix}\nabla G(x^k)^\top\alpha^k\\ 0\\ 0\\ 0\\ 0\end{pmatrix}+\begin{pmatrix}\nabla_x g(x^k,y^k)^\top\beta^k\\ \nabla_y g(x^k,y^k)^\top\beta^k\\ 0\\ \beta^k\\ 0\end{pmatrix}$$
$$+\begin{pmatrix}\nabla_x(\nabla_y L(x^k,y^k,u^k))^\top(\gamma_2^k-\gamma_1^k)\\ \nabla_y(\nabla_y L(x^k,y^k,u^k))^\top(\gamma_2^k-\gamma_1^k)\\ \nabla_y g(x^k,y^k)(\gamma_2^k-\gamma_1^k)\\ 0\\ 0\end{pmatrix}+\begin{pmatrix}0\\ 0\\ -\delta^k\\ 0\\ \delta^k\end{pmatrix}+\begin{pmatrix}0\\ 0\\ 0\\ a^{*k}\\ b^{*k}\end{pmatrix}=0,$$
$$\begin{aligned}&(\alpha^k)^\top G(x^k)=0 \text{ and } \forall j\in K(x^k):\ \alpha_j^k\geq 0,\\ &\gamma_{1,i}^k\geq 0,\ \gamma_{1,i}^k(\nabla_{y_i}L(x^k,y^k,u^k)+\varepsilon_1)=0,\ i=1,\dots,m\\ &\gamma_{2,i}^k\geq 0,\ \gamma_{2,i}^k(\nabla_{y_i}L(x^k,y^k,u^k)-\varepsilon_1)=0,\ i=1,\dots,m,\end{aligned}\tag{17}$$

see Mordukhovich [29, Theorem 5.21(iii)]. $N_A(a)$ denotes the limiting normal cone at a with respect to $A\subseteq\mathbb{R}^n$ and is empty if $a\notin A$, see [30].

By the definition of the normal cone,

$$N_{\Omega_{\varepsilon_2^k}}(a^k,b^k)=N_{\Omega_{\varepsilon_2^k,1}}(a_1^k,b_1^k)\times\cdots\times N_{\Omega_{\varepsilon_2^k,p}}(a_p^k,b_p^k),$$

cf. Outrata [31, Lemma 2.2]. Here,

$$\Omega_{\varepsilon_2,i}:=\{(a_i,b_i)\in\mathbb{R}\times\mathbb{R}: a_i\geq 0,\ b_i\geq 0,\ a_ib_i\leq\varepsilon_2\},\quad i=1,\dots,p.$$

Thus, $a_i^{*k}=-\beta_i^k=0$ for $a_i^k=-g_i(x^k,y^k)>0$, $b_i^k=u_i^k=0$ $(i\in I_u(x^k,y^k,u^k))$ and $b_i^{*k}=-\delta_i^k=0$ provided that $b_i^k=u_i^k>0$, $a_i^k=g_i(x^k,y^k)=0$ $(i\in I_g(x^k,y^k,u^k))$. Furthermore, $a_i^{*k}=-\beta_i^k\neq 0$ means that $i\in I_g(x^k,y^k,u^k)\cup I_{gu}(x^k,y^k,u^k,\varepsilon^k)$, and from $b_i^{*k}=-\delta_i^k\neq 0$, it follows that $i\in I_u(x^k,y^k,u^k)\cup I_{gu}(x^k,y^k,u^k,\varepsilon^k)$. Hence, for $\chi^k=0$, the first three equations in (17) violate MPEC-MFCQ. Thus, $\chi^k\neq 0$ and we can assume without loss of generality that $\chi^k=1$.

Let us show that the sequence of multipliers is bounded. Therefore, assume the opposite, i.e. there exists a subsequence

$$\{(\alpha^{k'},\beta^{k'},\gamma_1^{k'},\gamma_2^{k'},\delta^{k'},a^{*k'},b^{*k'})\}_{k'=1}^{\infty},\ (a^{*k'},b^{*k'})\in N_{\Omega_{\varepsilon_2^{k'}}}(a^{k'},b^{k'})$$

such that

$$\frac{(\alpha^{k'},\beta^{k'},\gamma_1^{k'},\gamma_2^{k'},\delta^{k'},a^{*k'},b^{*k'})}{\left\|(\alpha^{k'},\beta^{k'},\gamma_1^{k'},\gamma_2^{k'},\delta^{k'},a^{*k'},b^{*k'})\right\|} \overset{k'\to\infty}{\longrightarrow} (\alpha,\beta,\gamma_1,\gamma_2,\delta,a^*,b^*)$$
$$\neq (0,0,0,0,0,0,0).$$
$$\frac{1}{\left\|(\alpha^{k'},\beta^{k'},\gamma_1^{k'},\gamma_2^{k'},\delta^{k'},a^{*k'},b^{*k'})\right\|}(a^{*k'},b^{*k'}) \in N_{\Omega_{\varepsilon_2^{k'}}}(a^{k'},b^{k'})\ \forall\, k'$$

implies

$$(a^*,b^*) \in N_{\Omega_0}(a,b) \cup \left\{(w_1^*,w_2^*) \in \mathbb{R}^p \times \mathbb{R}^p : w_1^*, w_2^* > 0\right\}.$$

Using the results above, we deduce that

$$\begin{pmatrix} \nabla G(\bar{x})^\top \alpha \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \nabla_x g(\bar{x},\bar{y})^\top \beta \\ \nabla_y g(\bar{x},\bar{y})^\top \beta \\ 0 \\ \beta \\ 0 \end{pmatrix}$$
$$+ \begin{pmatrix} \nabla_x(\nabla_y L(\bar{x},\bar{y},\bar{u}))^\top(\gamma_2-\gamma_1) \\ \nabla_y(\nabla_y L(\bar{x},\bar{y},\bar{u}))^\top(\gamma_2-\gamma_1) \\ \nabla_y g(\bar{x},\bar{y})(\gamma_2-\gamma_1) \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ -\delta \\ 0 \\ \delta \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ a^* \\ b^* \end{pmatrix} = 0 \qquad (18)$$

holds for $\alpha \geq 0, \gamma_1 \geq 0, \gamma_2 \geq 0$, where

$$a_i^* = -\beta_i \neq 0 \Rightarrow i \in I_g(\bar{x},\bar{y},\bar{u}) \cup I_{gu}(\bar{x},\bar{y},\bar{u}),$$
$$b_i^* = \delta_i \neq 0 \Rightarrow i \in I_u(\bar{x},\bar{y},\bar{u}) \cup I_{gu}(\bar{x},\bar{y},\bar{u})$$

are fulfilled as well. The first three equations in (18) again violate MPEC-MFCQ. Summing up, $\alpha \geq 0$ with $\alpha^\top G(\bar{x}) = 0$, β, $\gamma := \gamma_2 - \gamma_1$ and δ satisfy (11). Note that γ is not bounded in sign.

Next, we show that the conditions (13)–(16) are valid. The set Ω_{ε_2}, together with an indication of the normals at different points, is plotted in Figure 1. If $\beta_i \neq 0$, then $i \in I_g(\bar{x},\bar{y},\bar{u}) \cup I_{gu}(\bar{x},\bar{y},\bar{u})$, which implies (13). If $\delta_i \neq 0$, then $i \in I_u(\bar{x},\bar{y},\bar{u}) \cup I_{gu}(\bar{x},\bar{y},\bar{u})$, which means that (14) is true. If $\bar{u}_i = g_i(\bar{x},\bar{y}) = 0$, there are four possible cases for sufficiently large k: $u_i^k = g_i(x^k,y^k) = 0$, $-u_i^k g_i(x^k,y^k) = \varepsilon_2^k$, $u_i^k > 0 \wedge g_i(x^k,y^k) = 0$ or $u_i^k = 0 \wedge g_i(x^k,y^k) < 0$. Then, $(a_i^{*k}, b_i^{*k}) \in \mathbb{R}^2_-$, $(a_i^{*k}, b_i^{*k}) \in \mathbb{R}^2_+$, $a_i^{*k} = 0 \ \wedge\ b_i^{*k} \leq 0$ or $a_i^{*k} \leq 0 \ \wedge\ b_i^{*k} = 0$, see Figure 1. For k tending to infinity, we derive $a_i^* b_i^* \geq 0$ or $\beta_i\delta_i \geq 0$ for all $i \in I_{gu}(\bar{x},\bar{y},\bar{u})$. By setting $\delta := \nabla_y g(\bar{x},\bar{y})\gamma$, this implies $\beta_i\delta_i \geq 0$ for $i \in I_{gu}(\bar{x},\bar{y},\bar{u})$ and, hence, C-stationarity.

■

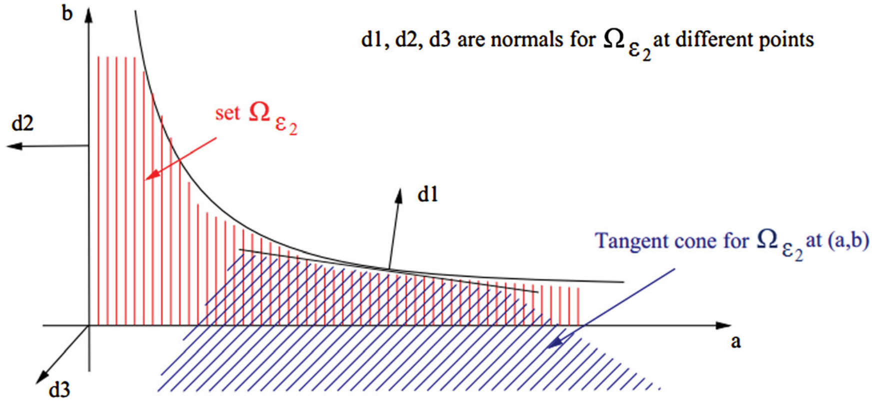

Figure 1. The set Ω_{ε_2}.

5. Some numerical tests

The use of problem (7) for solving the bilevel optimization problem (1), (2) has been numerically tested for the first time in the PhD thesis [19] of Mersha. Solving 23 (small) instances known from different sources, the best known solutions have been obtained for all but three instances. Since the bilevel optimization problem is a nonconvex optimization problem, it is not surprising that algorithms which are not especially tailored to compute global optima will not find global optima in general. Comparing this with the results obtained using model (4), it becomes obvious that using (7) is superior.

In his master's thesis [32], Rog compared three approaches for computing local optimal solutions of linear bilevel optimization problems. The first and second one are methods for computing sequences of (local) optimal solutions of the problems (4) and (7) for decreasing $\varepsilon_i, i = 1, 2$, the third one is the algorithm developed in [33] which goes back to the method in [34] and realizes a descent algorithm. Saboiev's method in [33] is related to the algorithm realized in [19] in the sense that it also uses an active set strategy for replacing the complementarity slackness constraint in the KKT conditions for the lower level problem. This results in a significantly higher computational effort of Saboiev's method, especially for instances with a larger set of so-called biactive indices.

Rog implemented in [32] all three methods using AMPL [35] and compared them solving a larger number of instances of linear bilevel optimization problems with random data. The essential results are the following:

(1) The algorithm using the model (7) has been implemented using different norms. Best results both with respect to small computational effort and large number of instances for which the best local optimal solution has been computed has been obtained using the Chebyshev norm.
(2) In about 40 % of the solved instances the best local optimal solution has been obtained using Mersha's approach as well as using Saboiev's algorithm.

The solution computed with the Scholtes' method was the best one in about 20 % of the instances.

(3) The computational effort of using (4) and (10) has been shown to be almost similar which is not surprising for this class of problems.

Summing up, an algorithm computing a sequence of local optimal solutions of the problem (10) for $\varepsilon_i,\ i = 1, 2$, tending to zero from above can be recommended for solving the bilevel optimization problem. The algorithm has a similar computational effort as the robust relaxation method [18] of Scholtes but can be shown to converge to a local optimal solution of the bilevel optimization problem (1), (2) under weak assumptions. Even if the approach in [17] converges to a local optimal solution of the bilevel problem, the suggested algorithm in this article often computes a better one.

6. Conclusion

For a long time, the KKT transformation (3) of the bilevel optimization problem (1), (2) has been used to solve this problem. Then, Dempe and Dutta [13] have shown that the local solutions of the nonconvex problem (3) do not need to be related to local optima of the bilevel optimization problem (1), (2). Hence, the use of the optimal value transformation was chosen to solve the problem. This is a nonsmooth optimization problem even under restrictive assumptions. In this article we found a simple approximation of the KKT transformation which enables us to apply general solution algorithms for nonconvex optimization problems in order to approximate the local optimal solution of the bilevel optimization problem which surely makes it possible to solve problems of higher dimension.

The need to demand additional assumptions for the lower level problem is not surprising if local optimal solutions are computed for the auxiliary problems. This was already observed in the application of decomposition approaches for solving optimization problems [36, 37].

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Research of both authors has been supported by Deutsche Forschungsgemeinschaft [grant number DE 650/7-1].

ORCID

S. Dempe http://orcid.org/0000-0001-6344-5152

References

[1] Bard JF. Practical bilevel optimization: algorithms and applications. Dordrecht: Kluwer Academic; 1998.

[2] Dempe S. Foundations of bilevel programming. Dordrecht: Kluwer Academic; 2002.

[3] Dempe S, Kalashnikov V, Pérez-Valdés GA, Kalashnykova N. Bilevel programming problems: theory, algorithms and application to energy networks. New York: Springer-Verlag; 2015.

[4] Dempe S. Annotated bibliography on bilevel programming and mathematical programs with equilibrium constraints. Optimization. 2003;52:333–359.

[5] Ye JJ, Zhu DL. Optimality conditions for bilevel programming problems. Optimization. 1995;33:9–27, with correction in Optimization 1997;39:361–366.

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

[7] Dempe S, Franke S. Solution algorithm for an optimistic linear Stackelberg problem. Comput Oper Res. 2014;41:277–281.

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

[9] Rockafellar RT, Wets RJ-B. Variational analysis. Berlin: Springer Verlag; 1998.

[10] Zemkoho AB. Bilevel programming: reformulations, regularity, and stationarity [Ph.D. thesis]. TU Bergakademie Freiberg; 2012.

[11] Mirrlees JA. The theory of moral hazard and unobservable behaviour: part I. Rev Econ Stud. 1999;66:3–21.

[12] Allende GB, Still G. Solving bilevel programs with the KKT-approach. Math Program. 2013;138:309–332.

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

[14] Luo Z-Q, Pang J-S, Ralph D. Mathematical programs with equilibrium constraints. Cambridge: Cambridge University Press; 1996.

[15] Outrata J, Kočvara M, Zowe J. Nonsmooth approach to optimization problems with equilibrium constraints. Dordrecht: Kluwer Academic; 1998.

[16] Scheel H, Scholtes S. Mathematical programs with equilibrium constraints: stationarity, optimality, and sensitivity. Math Oper Res. 2000;25:1–22.

[17] Scholtes S. Convergence properties of a regularization scheme for mathematical programs with complementarity constraints. SIAM J. Opt. 2001;11:918–936.

[18] Hoheisel T, Kanzow C, Schwartz A. Theoretical and numerical comparison of relaxation methods for mathematical programs with complementarity constraints. Math Program. 2013;137(1–2):257–288.

[19] Mersha AG. Solution methods for bilevel programming problems [Ph.D. thesis]. TU Bergakademie Freiberg; 2008.

[20] Mersha AG, Dempe S. Direct search algorithm for bilevel programming problems. Comput Optim Appl. 2011;49(1):1–15.

[21] Kojima M. Strongly stable stationary solutions in nonlinear programs. In: Robinson SM, editor. Analysis and computation of fixed points. New York: Academic Press; 1980. p. 93–138.

[22] Ralph D, Dempe S. Directional derivatives of the solution of a parametric nonlinear program. Math Program. 1995;70:159–172.

[23] Lignola MB, Morgan J. Existence of solutions to generalized bilevel programming problem. In: Migdalas A, Pardalos PM, and Värbrand P, editors. Multilevel optimization: algorithms and applications. Dordrecht: Kluwer Academic; 1998. p. 315–332.

[24] Lampariello L, Sagratella S. A bridge between bilevel programs and Nash games. J Optim Theory Appl. 2017;174(2):613–635.
[25] 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.
[26] Mersha AG, Dempe S. Feasible direction method for bilevel programming problem. Optimization. 2012;61(4–6):597–616.
[27] Gauvin J. A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming. Math Program. 1977;12:136–138.
[28] Bank B, Guddat J, Klatte D, Kummer B, Tammer K. Non-linear parametric optimization. Basel: Birkhäuser Verlag; 1983.
[29] Mordukhovich BS. Variational analysis and generalized differentiation, vol. 2: applications. Berlin: Springer Verlag; 2006.
[30] Mordukhovich BS. Variational analysis and generalized differentiation I: basic theory. Berlin: Springer; 2006.
[31] Outrata JV. Optimality conditions for a class of mathematical programs with equilibrium constraints. Math Oper Res. 1999;24:627–644.
[32] Rog R. Lösungsalgorithmen für die KKT-Transformation von Zwei-Ebenen-Optimierungsaufgaben [Master's thesis]. TU Bergakademie Freiberg, Fakultät für Mathematik und Informatik; 2017.
[33] Saboiev R. Solution methods for linear bilevel optimization problems [Ph.D. thesis]. TU Bergakademie Freiberg; 2016.
[34] Dempe S. A simple algorithm for the linear bilevel programming problem. Optimization. 1987;18:373–385.
[35] Fourer R, Gay DM, Kernighan BW. Ampl: a modeling language for mathematical programming. 1st ed. San Francisco: Scientific Press; 1993.
[36] De Miguel A-V. Two decomposition algorithms for nonconvex optimization problems with global variables [Ph.D. thesis]. Stanford University; 2001.
[37] Tammer K. The application of parametric optimization to the foundation and realization of a generalized primal decomposition approach. In: Guddat J, et al., editors. Proc. internat. conference "Parametric optimization and related topics", Plaue (1985). Berlin: Akademie-Verlag; 1987.