Math. Program., Ser. B (2012) 136:65–89
DOI 10.1007/s10107-012-0551-x

FULL LENGTH PAPER

# Bilevel optimization: on the structure of the feasible set

**H. Th. Jongen · V. Shikhman**

Received: 6 August 2010 / Accepted: 4 October 2011 / Published online: 18 May 2012


**Abstract** We consider bilevel optimization from the optimistic point of view. Let the pair $(x, y)$ denote the variables. The main difficulty in studying such problems lies in the fact that the lower level contains a global constraint. In fact, a point $(x, y)$ is feasible if $y$ solves a parametric optimization problem $L(x)$. In this paper we restrict ourselves to the special case that the variable $x$ is one-dimensional. We describe the generic structure of the feasible set $M$. Moreover, we discuss local reductions of the bilevel problem as well as corresponding optimality criteria. Finally, we point out typical problems that appear when trying to extend the ideas to higher dimensional $x$-dimensions. This will clarify the high intrinsic complexity of the general generic structure of the feasible set $M$ and corresponding optimality conditions for the bilevel problem $U$.

**Keywords** Bilevel programming · Parametric optimization · Structure of the feasible set · Local reduction · Optimality criteria

**Mathematics Subject Classification** 90C33

## 1 Introduction

We consider bilevel optimization problems as hierarchical problems of two decision makers, the so-called leader and follower. The follower selects his decision knowing the choice of the leader, whereas the latter has to anticipate the follower's

H. Th. Jongen · V. Shikhman (✉)
Department of Mathematics—C, RWTH Aachen University, 52056 Aachen, Germany
e-mail: shikhman@mathc.rwth-aachen.de

H. Th. Jongen
e-mail: jongen@rwth-aachen.de

response in his decision. Bilevel programming problems have been studied in the monographs [2] and [6]. We model the bilevel optimization problem in the so-called optimistic formulation. To this aim, assume that the follower solves the parametric optimization problem (lower level problem $L$)

$$L(x): \quad \min_{y} g(x, y) \quad \text{s.t.} \quad h_j(x, y) \geq 0, \quad j \in J \tag{1}$$

and that the leader's optimization problem (upper level problem $U$) is the following

$$U: \quad \min_{(x,y)} f(x, y) \quad \text{s.t.} \quad y \in \text{Argmin } L(x). \tag{2}$$

Above we have $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$ and the real valued mappings $f, g, h_j, j \in J$ belong to $C^3(\mathbb{R}^n \times \mathbb{R}^m)$, the space of three times continuously differentiable mappings. Argmin $L(x)$ denotes the solution set of the optimization problem $L(x)$. For simplicity, additional (in)equality constraints in defining $U$ are omitted.

The main goal of this article is to describe the generic structure of the bilevel feasible set $M$, where

$$M := \{(x, y) \mid y \in \text{Argmin } L(x)\}.$$

The special case with unconstrained one-dimensional lower level (i.e. $J = \emptyset$ and $m = 1$) is treated in [7]. In the latter paper the classification of 1-dimensional singularities was heavily used and for the higher dimensional case (i.e. $m > 1$) it is conjectured that a similar result will hold.

However, the situation becomes extremely difficult to describe if inequality constraints are present in the lower level (i.e. $J \neq \emptyset$). In particular, kinks and ridges will appear in the feasible set and such subsets might attract stable solutions of the bilevel problem. A simple example was presented in [7]. In this paper we restrict ourselves to the simplest case that the $x$-dimension is equal to one (i.e. $n = 1$), but no restrictions on the $y$-dimension. Then, the lower level $L(x)$ is a one-dimensional parametric optimization problem and we can exploit the well-known generic (five type) classification of so-called generalized critical points (cf. [13]) in order to describe the feasible set. Our main result (Theorems 4.1 and 4.2) states that—generically—the feasible set $M$ is the union of $C^2$ curves with boundary points and kinks which can be parametrized by means of the variable $x$. The appearance of the boundary points and kinks is due to certain degeneracies of the corresponding local solutions in the lower level as well as the change from local to global solutions. Outside of the latter points, the feasible points $(x, y(x)) \in M$ correspond to nondegenerate minimizers of the lower level $L(x)$. Although $dim(x) = 1$ might seem to be very restrictive, it should be noted that on typical curves in higher dimensional $x$-space the one-dimensional features as described in this paper will reappear on that curves.

The paper is organized as follows. In Sect. 2 we present some typical guiding examples. In Sect. 3 we recall the 5-Type classification of generalized critical points and discuss their possible appearance as global minimizers. In the latter case the connection with the local structure of the feasible set $M$ is pointed out. Section 4 contains the main results of our paper, i.e. the generic and stable structure of the feasible set $M$ as well as resulting optimality criteria for the bilevel problem $U$. In order to guarantee the existence of solutions of the lower level we will assume an appropriate compactness condition [cf. (24)].

In Sect. 5 we discuss some problems that appear when trying to extend the ideas to higher $x$-dimensions. On one hand we show that there are analytical aspects: it will not be possible to describe the feasible set $M$ at all points. This obstruction comes from classification of singularities. But, if we focus on a neighborhood of (local) solutions of the bilevel problem, then certain high order singularities can be avoided. On the other hand, there appear partitioning problems of combinatorial nature. This will clarify the high intrinsic complexity of the general generic structure of the feasible set $M$ and corresponding optimality criteria for the bilevel problem $U$. An interesting point for future research would be the discovery of a natural constraint qualification under which the whole feasible set $M$ might be expected to be a Lipschitz manifold with boundary.

Our notation is standard. The $n$-dimensional Euclidean space is denoted by $\mathbb{R}^n$. Given an arbitrary set $K \subset \mathbb{R}^n$ we denote its topological closure by $\overline{K}$. By $span\{a_1, \dots, a_t\}$ we denote the vector space over $\mathbb{R}$ generated by the finite number of vectors $a_1, \dots, a_t \in \mathbb{R}^n$ and $dim\{span\{a_1, \dots, a_t\}\}$ stands for its dimension. Given a differentiable function $F : \mathbb{R}^n \longrightarrow \mathbb{R}^n$, $DF$ denotes its Jacobian matrix. Given a differentiable function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, $Df$ denotes the row vector of partial derivatives of first order and $D^T f$ stands for the transposed vector.

Let $C^3(\mathbb{R}^n)$ denote the space of three times continuously differentiable real-valued functions. Let $C^3(\mathbb{R}^n)$ be endowed with the strong (or Whitney) $C^3$-topology, denoted by $C^3_s$ (cf. [10,14]). The $C^3_s$-topology is generated by allowing perturbations of the functions and their derivatives up to third order which are controlled by means of continuous positive functions. The product space of continuously differentiable functions will be topologized with the corresponding product topology. Note that the space of continuously differentiable functions endowed with the strong $C^3_s$-topology constitutes a Baire space. We say that a set is $C^3_s$-generic if it contains a countable intersection of $C^3_s$-open and $C^3_s$-dense subsets. Generic sets in a Baire space are dense as well.

## 2 Guiding examples

In this section we present several typical examples. They motivate our results on the structure of the bilevel feasible set $M$. In all examples the origin $0_{1+m}$ solves the bilevel problem $U$. Each example exhibits some kind of degeneracy in the lower level $L(x)$. Recall that $dim(x) = 1$ throughout the paper.

*Example 2.1*

$$f(x, y) := -x + 2y_1 + \varphi(y_2, \dots, y_m) \quad \text{with } \varphi \in C^3(\mathbb{R}^{m-1}, \mathbb{R}),$$
$$g(x, y) := (x - y_1)^2 + \sum_{j=2}^{m} y_j^2, \quad J = \{1\} \quad \text{and} \quad h_1(x, y) := y_1.$$

The degeneracy in the lower level $L(x)$ is the lack of strict complementarity at the origin $0_m$.

The bilevel feasible set $M$ becomes:

$$M = \{(x, \max(x, 0), 0, \ldots, 0) \mid x \in \mathbb{R}\}.$$

This example refers to Type 2 in the classification of Sect. 3. □

*Example 2.2*

$$f(x, y) := x + \sum_{j=1}^{m} y_j, \; g(x, y) := -y_1,$$
$$J = \{1\}, \quad h_1(x, y) := x - \sum_{j=1}^{m} y_j^2.$$

The degeneracy in the lower level $L(x)$ is the violation of the so-called Mangasarian–Fromovitz Constraint Qualification (MFCQ) (see Sect. 3) at the origin $0_m$. Moreover, the minimizer $0_m$ is a so-called Fritz–John point, but not a Karush–Kuhn–Tucker (KKT)-point.

The bilevel feasible set $M$ is a (half-)parabola:

$$M = \left\{(x, \sqrt{x}, 0, \ldots, 0) \mid x \geq 0\right\}.$$

This example refers to Type 4 in the classification of Sect. 3. □

*Example 2.3*

$$f(x, y) := x + \sum_{j=1}^{m} y_j, \quad g(x, y) := \sum_{j=1}^{m} y_j, \quad J = \{1, \ldots, m, m+1\},$$
$$h_j(x, y) := y_j, \quad j = 1, \ldots, m, \quad h_{m+1}(x, y) = x - \sum_{j=1}^{m} y_j.$$

The degeneracy in $L(0)$ is again the violation of the MFCQ at the origin $0_m$. However, in contrast to Example 2.2, the minimizer $0_m$ is a KKT-point now.

The bilevel feasible set $M$ becomes:

$$M = \{(x, 0, \ldots, 0) \mid x \geq 0\}.$$

This example refers to Type 5-1 in the classification of Sect. 3. □

*Example 2.4*

$$f(x, y) := -x + 2\sum_{j=1}^{m} y_j, \quad g(x, y) := \sum_{j=1}^{m} j y_j, \quad J = \{1, \ldots, m, m+1\},$$
$$h_j(x, y) := y_j, \quad j = 1, \ldots, m, \quad h_{m+1}(x, y) = -x + \sum_{j=1}^{m} y_j.$$

The degeneracy in $L(0)$ is the violation of the so-called linear independence constraint qualification (LICQ) at the origin $0_m$, whereas MFCQ is satisfied.

The bilevel feasible set $M$ becomes:

$$M = \{(x, \max(x, 0), 0, \dots, 0) \mid x \in \mathbb{R}\}.$$

This example refers to Type 5-2 in the classification of Sect. 3. □

Note that the feasible set $M$ exhibits a kink in Examples 2.1, 2.4, whereas it has a boundary in Examples 2.2, 2.3. Moreover, the minimizer $0_m$ in $L(0)$ is strongly stable (in the terminology of Kojima [11]) in Examples 2.1, 2.4, but not in Examples 2.2, 2.3.

We note that, despite of degeneracies in the lower level, the structure of the bilevel feasible set $M$ with its kinks and boundaries remains stable under small $C^3_s$-perturbations of the defining functions.

## 3 Five types classification and global minimizers

We consider the lower level problem $L(\cdot)$ in a one-dimensional parametric optimization setting, i.e. $dim(x) = 1$:

$$L(x): \quad \min_y g(x, y) \quad \text{s.t.} \quad h_j(x, y) \geq 0, \quad j \in J.$$

We denote its feasible set by

$$M(x) := \left\{ y \in \mathbb{R}^m \mid h_j(x, y) \geq 0, \ j \in J \right\}$$

and for $\bar{y} \in M(\bar{x})$ the active index set by

$$J_0(\bar{x}, \bar{y}) := \left\{ j \in J \mid h_j(\bar{x}, \bar{y}) = 0 \right\}.$$

**Definition 3.1** (*Generalized critical point*) A point $\bar{y} \in M(\bar{x})$ is called a generalized critical point (g.c. point) for $L(\bar{x})$ if the set of vectors

$$\left\{ D_y g(\bar{x}, \bar{y}), \ D_y h_j(\bar{x}, \bar{y}), \ j \in J_0(\bar{x}, \bar{y}) \right\} \tag{3}$$

is linearly dependent.

The critical set for $L(\cdot)$ is given by

$$\Sigma := \left\{ (x, y) \in \mathbb{R}^{1+m} \mid y \text{ is g.c. point for } L(x) \right\}.$$

In [13] it is shown that generically each point of $\Sigma$ is one of the Types 1–5. In what follows, we shortly recall the Types 1–5 and consider the structure of $\Sigma$ locally around particular g.c. points being local minimizers for $L(\cdot)$. Here, we focus on such parts of $\Sigma$ which correspond to (local) minimizers, i.e.

$$\Sigma_{min} := \{(x, y) \in \Sigma \mid y \text{ is a local minimizer for } L(x)\}$$

in a neighborhood of $(\bar{x}, \bar{y}) \in \Sigma_{min}$. We refer to [12] for the indication of the latter issue.

### 3.1 Points of Type 1

A point $(\bar{x}, \bar{y}) \in \Sigma$ is of Type 1 if $\bar{y}$ is a nondegenerate critical point for $L(\bar{x})$. It means that the following conditions ND1–ND3 hold.

ND1: Linear independence constraint qualification (LICQ) is satisfied at $(\bar{x}, \bar{y})$, i.e. the set of vectors

$$\left\{D_y h_j(\bar{x}, \bar{y}),\ j \in J_0(\bar{x}, \bar{y})\right\} \tag{4}$$

is linearly independent.
From (3) and (4) we see that there exist (Lagrange multipliers) $\bar{\mu}_j, \quad j \in J_0(\bar{x}, \bar{y})$, such that

$$D_y g(\bar{x}, \bar{y}) = \sum_{j \in J_0(\bar{x}, \bar{y})} \bar{\mu}_j D_y h_j(\bar{x}, \bar{y}). \tag{5}$$

ND2: $\bar{\mu}_j \neq 0, \quad j \in J_0(\bar{x}, \bar{y})$,
ND3: $D^2_{yy} L(\bar{x}, \bar{y})_{|T_{\bar{y}} M(\bar{x})}$ is nonsingular.

Here, the matrix $D^2_{yy} L(\bar{x}, \bar{y})$ stands for the Hessian w.r.t. $y$ variables of the Lagrange function $L$,

$$L(x, y) := g(x, y) - \sum_{j \in J_0(\bar{x}, \bar{y})} \bar{\mu}_j h_j(x, y). \tag{6}$$

and $T_{\bar{y}} M(\bar{x})$ denotes the tangent space of $M(\bar{x})$ at $\bar{y}$,

$$T_{\bar{y}} M(\bar{x}) := \left\{\xi \in \mathbb{R}^m \mid D_y h_j(\bar{x}, \bar{y}) \cdot \xi = 0,\ j \in J_0(\bar{x}, \bar{y})\right\}. \tag{7}$$

Condition ND3 means that the matrix $V^T D^2_{yy} L(\bar{x}, \bar{y}) V$ is nonsingular, where $V$ is some matrix whose columns form a basis for the tangent space $T_{\bar{y}} M(\bar{x})$.

The linear index LI, resp. linear coindex LCI, is defined to be the number of $\bar{\mu}_j$ in (5) which are negative, resp. positive. The quadratic index QI, resp. quadratic co-index QCI, is defined to be the number of negative, resp. positive eigenvalues of $D^2_{yy} L(\bar{x}, \bar{y})_{|T_{\bar{y}} M(\bar{x})}$.

*Characteristic numbers: LI, LCI, QI, QCI*

It is well-known that conditions ND1–ND3 allow us to apply the implicit function theorem and obtain unique $C^2$-mappings $y(x)$, $\mu_j(x)$, $j \in J_0(\bar{x}, \bar{y})$ in an open neighborhood of $\bar{x}$. It holds: $y(\bar{x}) = \bar{y}$ and $\mu_j(\bar{x}) = \bar{\mu}_j$, $j \in J_0(\bar{x}, \bar{y})$, moreover, for $x$ sufficiently close to $\bar{x}$ the point $y(x)$ is a nondegenerate critical point for $L(x)$ with Lagrange multipliers $\mu_j(x)$, $\quad j \in J_0(\bar{x}, \bar{y})$ having the same indices LI, LCI, QI, QCI as $\bar{y}$. Hence, locally around $(\bar{x}, \bar{y})$ we can parametrize the set $\Sigma$ by means of a unique $C^2$-map $x \mapsto (x, y(x))$. If $\bar{y}$ is additionally a local minimizer for $L(\bar{x})$, i.e. LI = QI = 0, then we get locally around $(\bar{x}, \bar{y})$:

$$\Sigma_{min} = \{(x, y(x)) \mid x \text{ sufficiently close to } \bar{x}\}.$$

## 3.2 Points of Type 2

A point $(\bar{x}, \bar{y}) \in \Sigma$ is of Type 2 if the following conditions A1–A6 hold:

A1: LICQ is satisfied at $(\bar{x}, \bar{y})$

A2: $J_0(\bar{x}, \bar{y}) \neq \emptyset$

After renumbering we may assume that $J_0(\bar{x}, \bar{y}) = \{1, \ldots, p\}$, $p \geq 1$. Then, we have

$$D_y g(\bar{x}, \bar{y}) = \sum_{j=1}^{p} \bar{\mu}_j D_y h_j(\bar{x}, \bar{y}). \tag{8}$$

A3: In (8) exactly one of the Lagrange multipliers vanishes.

After renumbering we may assume that $\bar{\mu}_p = 0$ and $\bar{\mu}_j \neq 0$, $j = 1, \ldots, p-1$. Let $L$ and $T_{\bar{y}}M(\bar{x})$ be defined as in (6) and (7), respectively.

A4: $D^2_{yy}L(\bar{x}, \bar{y})_{|T_{\bar{y}}M(\bar{x})}$ is nonsingular

We set

$$T^+_{\bar{y}}M(\bar{x}) := \left\{\xi \in \mathbb{R}^m \mid D_y h_j(\bar{x}, \bar{y}) \cdot \xi = 0,\ j \in J_0(\bar{x}, \bar{y}) \backslash \{p\}\right\}.$$

A5: $D^2_{yy}L(\bar{x}, \bar{y})_{|T^+_{\bar{y}}M(\bar{x})}$ is nonsingular

Let $W$ be a matrix with $m$ rows, whose columns form a basis of the linear space $T^+_{\bar{y}}M(\bar{x})$. Put $\Phi = (h_1, \ldots, h_{p-1})^T$ and define the $m \times 1$-vectors:

$$\alpha := -\left[\left(D_y\Phi \cdot D_y^T\Phi\right)^{-1} \cdot D_y\Phi\right]^T \cdot D_x\Phi,$$

$$\beta = -W \cdot \left(W^T \cdot D^2_{yy}L \cdot W\right)^{-1} \cdot W^T \left[D^2_{yy}L \cdot \alpha + D_x D_y^T L\right]$$

Note that all partial derivatives are evaluated at $(\bar{x}, \bar{y})$. Next, we put

$$\gamma := D_x h_p(\bar{x}, \bar{y}) + D_y h_p(\bar{x}, \bar{y})(\alpha + \beta).$$

A6: $\gamma \neq 0$

Let $\delta_1$ and $\delta_2$ denote the number of negative eigenvalues of $D^2_{yy}L(\bar{x}, \bar{y})_{|T^+_{\bar{y}}M(\bar{x})}$ and $D^2_{yy}L(\bar{x}, \bar{y})_{|T_{\bar{y}}M(\bar{x})}$, respectively, and put $\delta := \delta_1 - \delta_2$.

*Characteristic numbers: $sign(\gamma)$, $\delta$*

We proceed with the local analysis of the set $\Sigma$ in a neighborhood of $(\bar{x}, \bar{y})$.

(a) We consider the following associated optimization problem (without the $p$th constraint):

$$\widetilde{L}(x): \quad \underset{y \in \mathbb{R}^m}{\text{minimize}}\, g(x, y) \quad \text{s.t.} \quad h_j(x, y) \geq 0, \quad j \in J \backslash \{p\}. \tag{9}$$

It is easy to see that $\bar{y}$ is a nondegenerate critical point for $\widetilde{L}(\bar{x})$ due to A1, A3, A5. As in Sect. 3.1 we get a unique $C^2$-map $x \mapsto (x, \widetilde{y}(x))$. The latter curve belongs to $\Sigma$ as far as $\psi(x)$ is nonnegative, where

$$\psi(x) := h_p(x, \widetilde{y}(x)).$$

A few calculations show that

$$\frac{d\widetilde{y}(\bar{x})}{dx} = \alpha + \beta \quad \text{and, hence,} \quad \frac{d\psi(\bar{x})}{dx} = \gamma. \tag{10}$$

Consequently, if we walk along the curve $x \mapsto (x, \widetilde{y}(x))$ as $x$ increases, then at $x = \bar{x}$ we leave (enter) the feasible set $M(x)$ according to $sign(\gamma) = -1(+1)$ (cf. A6).

(b) We consider the following associated optimization problem (with the $p$th constraint as equality):

$$\widehat{L}(x): \quad \underset{y \in \mathbb{R}^m}{\text{minimize}}\, g(x, y) \quad \text{s.t.} \quad h_j(x, y) \geq 0, \quad j \in J, \quad h_p(x, y) = 0. \tag{11}$$

It is easy to see that $\bar{y}$ is a nondegenerate critical point for $\widehat{L}(\bar{x})$ due to A1, A3, A4. Using results of Sect. 3.1 we get a unique $C^2$-map $x \mapsto (x, \widehat{y}(x))$. Note that $h_p(x, \widehat{y}(x)) \equiv 0$. Moreover, it can be calculated that

$$sign(\gamma) \cdot sign\left(\frac{d\mu_p(\bar{x})}{dx}\right) = -1 \quad (\text{resp.}\ +1) \quad \text{iff} \quad \delta = 0 \quad (\text{resp.}\ \delta = 1). \tag{12}$$

Altogether, since the curve $x \mapsto (x, \widetilde{y}(x))$ traverses the zero set "$h_p = 0$" at $(\bar{x}, \bar{y})$ transversally (cf. A6), it follows that $x \mapsto (x, \widetilde{y}(x))$ and $x \mapsto (x, \widehat{y}(x))$ intersect at $(\bar{x}, \bar{y})$ under a nonvanishing angle. Obviously, in a neighborhood of $(\bar{x}, \bar{y})$ the set $\Sigma$ consists of $x \mapsto (x, \widehat{y}(x))$ and that part of $x \mapsto (x, \widetilde{y}(x))$ on which $h_p$ is nonnegative.

Let now additionally assume that $\bar{y}$ is a local minimizer for $L(\bar{x})$. Then, $\bar{\mu}_j > 0$, $j \in J_0(\bar{x}, \bar{y})\backslash\{p\}$ in A3, and the matrix $D^2_{yy}L(\bar{x}, \bar{y})_{|T_{\bar{y}}M(\bar{x})}$ is positive definite in A4, hence, $\delta_2 = 0$.

We consider two cases for $\delta = 0$ or $\delta = 1$.

*Case* $\delta = 0$*:*

In this case $D^2_{yy}L(\bar{x}, \bar{y})_{|T^+_{\bar{y}}M(\bar{x})}$ is positive definite in A5. Hence, $\bar{y}$ is a strongly stable local minimizer for $L(\bar{x})$ (see [11] for details on the strong stability). Moreover, $\widetilde{y}(x)$ is a local minimizer for $L(x)$ if $h_p(x, \widetilde{y}(x)) > 0$. Otherwise, $\widehat{y}(x)$ is a local minimizer for $L(x)$ since the corresponding Lagrange multiplier $\mu_p(x)$ becomes positive due to (12). Note that the sign of $h_p(x, \widetilde{y}(x))$ is corresponding to $sign(\gamma)$ as obtained in (10).

Then, we get locally around $(\bar{x}, \bar{y})$:

$$\Sigma_{min} = \left\{(x, y(x)) \mid y(x) := \left\{ \begin{array}{l} \widetilde{y}(x),\ x \le \bar{x} \\ \widehat{y}(x),\ \bar{x} \le x \end{array} \right\} \text{ if } sign(\gamma) = -1\right.$$

and

$$\Sigma_{min} = \left\{(x, y(x)) \mid y(x) := \left\{ \begin{array}{l} \widehat{y}(x),\ x \le \bar{x} \\ \widetilde{y}(x),\ \bar{x} \le x \end{array} \right\} \text{ if } sign(\gamma) = +1.\right.$$

*Case* $\delta = 1$*:*

In this case $D^2_{yy}L(\bar{x}, \bar{y})_{|T^+_{\bar{y}}M(\bar{x})}$ has exactly one negative eigenvalue. Thus, we obtain that the optimal value of the following optimization problem is negative:

$$\begin{aligned} &\underset{\xi \in \mathbb{R}^m}{\text{minimize}}\ \xi^T \cdot D^2_{yy}g(\bar{x}, \bar{y}) \cdot \xi \quad \text{s.t.} \quad \|\xi\| = 1, \quad \xi \in T^+_{\bar{y}}M(\bar{x}), \\ &D_y h_p(\bar{x}, \bar{y}) \cdot \xi \ge 0. \end{aligned}$$

In view of that, at $(\bar{x}, \bar{y})$ we can find a quadratic descent direction $\xi$ for $L(\bar{x})$. Thus, $\bar{y}$ is not a local minimizer for $L(\bar{x})$ which contradicts to the above assumption. We conclude that this case does not occur in $\Sigma_{min}$.

### 3.3 Points of Type 3

A point $(\bar{x}, \bar{y}) \in \Sigma$ is of Type 3 if the following conditions B1–B4 hold:

B1: LICQ is satisfied at $(\bar{x}, \bar{y})$

After renumbering we may assume in case $J_0(\bar{x}, \bar{y}) \neq \emptyset$ that $J_0(\bar{x}, \bar{y}) = \{1, \ldots, p\}$, $p \ge 1$. Then, we have

$$D_y g(\bar{x}, \bar{y}) = \sum_{j=1}^{p} \bar{\mu}_j D_y h_j(\bar{x}, \bar{y}). \tag{13}$$

B2: In (13) we have $\bar{\mu}_j \neq 0,\ j = 1, \ldots, p$.

Let $L$ and $T_{\bar{y}}M(\bar{x})$ be defined as in (6) and (7), respectively.

B3: Exactly one eigenvalue of $D^2_{yy}L(\bar{x}, \bar{y})_{|T_{\bar{y}}M(\bar{x})}$ vanishes.

Let $V$ be a matrix, whose columns form a basis for the tangent space $T_{\bar{y}}M(\bar{x})$. According to B3, let $w$ be a nonvanishing vector such that $V^T \cdot D^2_{yy}L(\bar{x}, \bar{y}) \cdot Vw = 0$, and put $v := V \cdot w$. Put $\Phi = (h_1, \ldots, h_{p-1})^T$ and define

$$\beta_1 := v^T(D^3_{yyy}L \cdot v)v - 3v^T D^2_{yy}L \cdot \left(\left(D_y\Phi \cdot D_y^T\Phi\right)^{-1} \cdot D_y\Phi\right) \cdot (v^T D^2_{yy}\Phi v),$$

$$\beta_2 := D_x(D_yL \cdot v) - D_x^T\Phi \cdot \left(\left(D_y\Phi \cdot D_y^T\Phi\right)^{-1} \cdot D_y\Phi\right) \cdot D^2_{yy}L \cdot v.$$

Note that all partial derivatives are evaluated at $(\bar{x}, \bar{y})$. Next, we put

$$\beta := \beta_1 \cdot \beta_2.$$

B4: $\beta \neq 0$

Let $\alpha$ denote the number of negative eigenvalues of $D^2_{yy}L(\bar{x}, \bar{y})_{|T_{\bar{y}}M(\bar{x})}$.

*Characteristic numbers: $sign(\beta)$, $\alpha$*

It turns out that in a neighborhood of $(\bar{x}, \bar{y})$ the set $\Sigma$ is a one-dimensional $C^2$−manifold. Moreover, the parameter $x$, viewed as a function on $\Sigma$, has a (nondegenerate) local maximum, resp. local minimizer, at $(\bar{x}, \bar{y})$ according to $sign(\beta) = +1$, resp. $sign(\beta) = -1$. Consequently, the set $\Sigma$ can be locally approximated by means of a parabola. In particular, if we approach the point $(\bar{x}, \bar{y})$ along $\Sigma$, the path of local minimizers (with QI $= \alpha = 0$) stops and the local minimizer switches into a saddlepoint (with QI $= \alpha + 1 = 1$). Moreover, note that at $(\bar{x}, \bar{y})$ there exists a unique (tangential) direction of cubic descent, hence, $\bar{y}$ can not be a local minimizer for $L(\bar{x})$. Hence, this case does not occur in $\Sigma_{min}$.

## 3.4 Points of Type 4

A point $(\bar{x}, \bar{y}) \in \Sigma$ is of Type 4 if the following conditions C1–C6 hold:

C1: $J_0(\bar{x}, \bar{y}) \neq \emptyset$
After renumbering we may assume that $J_0(\bar{x}, \bar{y}) = \{1, \ldots, p\}$, $p \geq 1$.
C2: $dim\left\{span\left\{D_yh_j(\bar{x}, \bar{y}), \quad j \in J_0(\bar{x}, \bar{y})\right\}\right\} = p - 1$
C3: $p - 1 < m$
From C2 we see that there exist $\bar{\mu}_j$, $j \in J_0(\bar{x}, \bar{y})$, not all vanishing such that

$$\sum_{j=1}^{p} \bar{\mu}_j D_yh_j(\bar{x}, \bar{y}) = 0. \tag{14}$$

Note that the numbers $\bar{\mu}_j$, $j \in J_0(\bar{x}, \bar{y})$ are unique up to a common multiple.
C4: $\bar{\mu}_j \neq 0, \quad j \in J_0(\bar{x}, \bar{y})$ and we normalize the $\bar{\mu}_j$'s by setting $\bar{\mu}_p = 1$

We define furthermore

$$L(x, y) := h_p(x, y) + \sum_{j=1}^{p-1} \bar{\mu}_j h_j(x, y) \quad \text{and}$$

$$T_{\bar{y}} M(\bar{x}) := \left\{ \xi \in \mathbb{R}^m \mid D_y h_j(\bar{x}, \bar{y}) \cdot \xi = 0,\ j \in J_0(\bar{x}, \bar{y}) \right\}$$

Let $W$ be a matrix, whose columns form a basis for $T_{\bar{y}} M(\bar{x})$. Define

$$A := D_x L \cdot W^T \cdot D^2_{yy} L \cdot W \quad \text{and} \quad w := W^T \cdot D^T_y g,$$

all partial derivatives being evaluated at $(\bar{x}, \bar{y})$.

C5: $A$ is nonsingular
Finally define

$$\alpha := w^T \cdot A^{-1} \cdot w.$$

C6: $\alpha \neq 0$

Let $\beta$ denote the number of positive eigenvalues of $A$. Let $\gamma$ be the number of negative $\bar{\mu}_j,\ j \in \{1, \ldots, p-1\}$ and put $\delta := D_x L(\bar{x}, \bar{y})$.

*Characteristic numbers:* $sign(\alpha)$, $sign(\delta)$, $\gamma$, $\beta$

We proceed with the local analysis of the set $\Sigma$ in a neighborhood of $(\bar{x}, \bar{y})$. Conditions C2, C4 and C5 imply that (locally around $(\bar{x}, \bar{y})$) at all points $(x, y) \in \Sigma$—apart from $(\bar{x}, \bar{y})$—LICQ holds. Moreover, the active set $J_0(\cdot)$ is locally constant $(= J_0(\bar{x}, \bar{y}))$ on $\Sigma$. Having these facts in mind, we consider the following map $\Psi : \mathbb{R} \times \mathbb{R}^m \times \mathbb{R}^{p-1} \times \mathbb{R} \longrightarrow \mathbb{R}^m \times R^p$:

$$\Psi(x, y, \mu, \lambda) := \begin{pmatrix} \lambda D_y g(x, y) + D_y h_p(x, y) + \sum_{j=1}^{p-1} \mu_j D_y h_j(x, y) \\ h_j(x, y) = 0, \quad j = 1, \ldots, p-1 \\ \lambda g(x, y) + h_p(x, t) + \sum_{j=1}^{p-1} \mu_j h_j(x, y) \end{pmatrix}.$$

Note that $\Psi(\bar{x}, \bar{y}, \bar{\mu}, 0) = 0$ and $D_{x,y,\mu}\Psi(\bar{x}, \bar{y}, \bar{\mu}, 0)$ is nonsingular due to C5 and C6. Hence, there exists the unique $C^2$-mapping $\lambda \mapsto (x(\lambda), y(\lambda), \mu(\lambda))$ such that $\Psi(x(\lambda), y(\lambda), \mu(\lambda), \lambda) \equiv 0$ and $(x(0), y(0), \mu(0)) = (\bar{x}, \bar{y}, \bar{\mu})$. Further, it is not hard to see that locally around $(\bar{x}, \bar{y})$

$$\Sigma = \{(x(\lambda), y(\lambda)) \mid \lambda \text{ sufficiently close to } 0\}.$$

The Lagrange multipliers corresponding to $(x(\lambda), y(\lambda))$ are

$$\left(-\frac{\mu_j(\lambda)}{\lambda},\ j = 1, \ldots, p-1, -\frac{1}{\lambda}\right). \tag{15}$$

It turns out that in a neighborhood of $(\bar{x}, \bar{y})$ the set $\Sigma$ is a one-dimensional $C^2-$manifold. The parameter $x$, viewed as a function on $\Sigma$, has a (nondegenerate) local maximum, resp. local minimizer, at $(\bar{x}, \bar{y})$ according to $sign(\alpha) = +1$, resp. $sign(\alpha) = -1$. Consequently, the set $\Sigma$ can be locally approximated by means of a parabola.

Let now additionally assume that $\bar{y}$ is a local minimizer for $L(\bar{x})$. Then, $\bar{\mu}_j > 0$, $j = 1, \ldots, p-1$ in C4 and, hence,

$$\gamma = 0. \tag{16}$$

Moreover, the matrix $W^T \cdot D^2_{yy}L \cdot W$ is negative definite. In particular, we get

$$\beta = \begin{cases} n-(p-1) & \text{if } \ sign(\delta) = -1, \\ 0 & \text{if } \ sign(\delta) = 1 \end{cases} \tag{17}$$

We are interested in the local structure of $\Sigma_{min}$ at $(\bar{x}, \bar{y})$. It is clear from (15) that $\lambda$ must be nonpositive if following the branch of local minimizers.

We consider two cases with respect to $sign(\alpha)$ and $sign(\delta)$

*Case 1:* $sign(\alpha) = sign(\delta)$

A few calculations show that

$$D_\lambda g(x(\lambda), y(\lambda))_{\lambda=0} = -\alpha \cdot \delta.$$

Hence, $D_\lambda g(x(\lambda), y(\lambda))_{\lambda=0} < 0$ and $g(x(\cdot), y(\cdot))$ is strictly decreasing when passing $\lambda = 0$. Consequently, the possible branch of local minimizers corresponding to $\lambda \le 0$ can not be one of global minimizers. We omit this case in view of our further interest in *global* minimizers in the context of bilevel programming problems.

*Case 2:* $sign(\alpha) \neq sign(\delta)$

In this case we get locally around $(\bar{x}, \bar{y})$:

$$\Sigma_{min} = \{(x(\lambda), y(\lambda)) \mid \lambda \le 0\}.$$

In fact, for $sign(\alpha) = 1$ and $sign(\delta) = -1$ the linear and quadratic indices of $y(\lambda)$ for $L(x(\lambda))$, $\lambda < 0$ are

$$\mathrm{LI} = \gamma = 0, \quad QI = n - p - \beta + 1 = n - p - (n - p + 1) + 1 = 0.$$

For $sign(\alpha) = -1$ and $sign(\delta) = 1$ the linear and quadratic indices of $y(\lambda)$ for $L(x(\lambda))$, $\lambda < 0$ are

$$\mathrm{LI} = \gamma = 0, \quad \mathrm{QI} = \beta = 0.$$

Confer (16) and (17) for the values of $\gamma$ and $\beta$, respectively.

### 3.5 Points of Type 5

A point $(\bar{x}, \bar{y}) \in \Sigma$ is of Type 5 if the following conditions D1–D4 hold:

D1: $|J_0(\bar{x}, \bar{y})| = m + 1$

D2: The set of vectors

$$\left\{Dh_j(\bar{x}, \bar{y}),\ j \in J_0(\bar{x}, \bar{y})\right\}$$

is linearly independent (derivatives in $\mathbb{R}^{m+1}$)

After renumbering we may assume that $J_0(\bar{x}, \bar{y}) = \{1, \ldots, p\}$, $p \geq 2$.

From D1, D2 we see that there exist $\mu_j$, $j \in J_0(\bar{x}, \bar{y})$, not all vanishing such that

$$\sum_{j=1}^{p} \mu_j D_y h_j(\bar{x}, \bar{y}) = 0. \tag{18}$$

Note that the numbers $\mu_j$, $j \in J_0(\bar{x}, \bar{y})$ are unique up to a common multiple.

D3: $\mu_j \neq 0$, $j \in J_0(\bar{x}, \bar{y})$

From D1, D2 it follows that there exist unique numbers $\beta_j$, $y \in J_0(\bar{x}, \bar{y})$ such that

$$Dg(\bar{x}, \bar{y}) = \sum_{j=1}^{p} \beta_j Dh_j(\bar{x}, \bar{y}). \tag{19}$$

Put

$$\Delta_{ij} := \beta_i - \beta_j \cdot \frac{\mu_i}{\mu_j} \quad \text{for } i, j = 1, \ldots, p$$

and let $\Delta$ be the $p \times p$ matrix with $\Delta_{ij}$ as its $(i, j)$th element.

D4: All off-diagonal elements of $\Delta$ do not vanish

We set

$$L(\bar{x}, \bar{y}) = \sum_{j=1}^{p} \mu_j h_j(\bar{x}, \bar{y}).$$

From D2 we see that $D_x L(\bar{x}, \bar{y}) \neq 0$. We define:

$$\gamma_j := sign\left(\mu_j \cdot D_x L(\bar{x}, \bar{y})\right) \quad \text{for } i, j = 1, \ldots, p.$$

By $\delta_j$ we denote the number of negative entries in the $j$th column of $\Delta$, $j = 1, \ldots, p$.

*Characteristic numbers:* $\gamma_j, \delta_j$, $j = 1, \ldots, p$

We proceed with the local analysis of the set $\Sigma$ in a neighborhood of $(\bar{x}, \bar{y})$. Conditions D1-D3 imply that locally around $(\bar{x}, \bar{y})$ at all points $(x, y) \in \Sigma \backslash \{(\bar{x}, \bar{y})\}$ LICQ holds. Combining (18) and (19) we obtain:

$$D_x g(\bar{x}, \bar{y}) = \sum_{j=1}^{p} \left( \beta_j - \beta_q \cdot \frac{\mu_j}{\mu_q} \right) D_x h_j(\bar{x}, \bar{y}), \quad q = 1, \ldots, p. \tag{20}$$

These both facts imply that for all $(x, y) \in \Sigma \backslash \{(\bar{x}, \bar{y})\}$ in a neighborhood of $(\bar{x}, \bar{y})$:

$$\| J_0(x, y) \| = m \quad \text{and} \quad J_0(x, y) = J_0(\bar{x}, \bar{y}) \backslash \{q\} \tag{21}$$

with some $q \in \{1, \ldots, p\}$ (in general, depending on $(x, y)$).

We put

$$M_q := \left\{ (x, y) \mid h_j(x, y) = 0, \quad j \in J_0(\bar{x}, \bar{y}) \backslash \{q\} \right\} \quad \text{and}$$
$$M_q^+ := \left\{ (x, y) \in M_q \mid h_q(x, y) \geq 0 \right\}.$$

From (20) and (21) it is easy to see that locally around $(\bar{x}, \bar{y})$

$$\Sigma = \bigcup_{q=1}^{p} M_q^+.$$

The indices (LI, LCI, QI, QCI) along $M_q^+ \backslash \{(\bar{x}, \bar{y})\}$ are equal $(\delta_q, m - \delta_q, 0, 0)$. Let $q \in \{1, \ldots, p\}$ be fixed. $M_q$ is a one-dimensional $C^3$-manifold due to D2. Since the set of vectors

$$\left\{ D_y h_j(\bar{x}, \bar{y}), \quad j \in J_0(\bar{x}, \bar{y}) \backslash \{q\} \right\}$$

is linearly independent, we can parametrize $M_q$ by means of the unique $C^3$-mapping $x \mapsto (x, y^q(x))$ with $y^q(\bar{x}) = \bar{y}$. A short calculation shows that

$$sign \left( \left. \frac{d h_q(x, y^q(x))}{dx} \right|_{x=\bar{x}} \right) = \gamma_q.$$

Hence, by increasing $x$, $M_q^+$ emanates from $(\bar{x}, \bar{y})$, resp. ends at $(\bar{x}, \bar{y})$ according to $\gamma_q = +1$, resp. $\gamma_q = -1$.

Let now additionally assume that $\bar{y}$ is a local minimizer for $L(\bar{x})$. For describing $\Sigma_{min}$ we define the so-called Karush–Kuhn–Tucker subset

$$\Sigma_{\text{KKT}} := cl \, \{(x, y) \in \Sigma \mid (x, y) \text{ is of Type 1 with LI} = 0\}.$$

It is shown in [13, Theorem 4.1] that—generically—$\Sigma_{\text{KKT}}$ is a one-dimensional (piecewise $C^2$-) manifold with boundary. In particular, $(x, y) \in \Sigma_{\text{KKT}}$ is a boundary point iff at $(x, y)$ we have: $J_0(x, y) \neq \emptyset$ and the Mangasarian–Fromovitz Constraint

Qualification (MFCQ) fails to hold. We recall that MFCQ is said to be satisfied for $(x, y)$, $y \in M(x)$, if there exists a vector $\xi \in \mathbb{R}^m$ such that

$$D_y h_j(x, y) \cdot \xi > 0 \quad \text{for all } j \in J_0(x, y).$$

Now we consider two cases with respect to the signs of $\mu_j$, $j \in J_0(\bar{x}, \bar{y})$:

*Case 1: all $\mu_j$, $j \in J_0(\bar{x}, \bar{y})$ have the same sign*

Recalling (18) we obtain that MFCQ is not fulfilled at $(\bar{x}, \bar{y})$. Hence, $(\bar{x}, \bar{y})$ is a boundary point of $\Sigma_{\text{KKT}}$. Having in mind the formulas for the indices (LI= $\delta_q$, LCI= $m - \delta_q$, QI=0, QCI=0) along $M_q^+ \backslash \{(\bar{x}, \bar{y})\}$ we obtain that $\delta_q = 0$ for some $q \in \{1, \ldots, p\}$. Moreover, a simple calculation shows

$$\Delta_{ij} = -\frac{\mu_i}{\mu_j} \cdot \Delta_{ji}, \quad i, j = 1, \ldots, p. \tag{22}$$

Since all $\mu_j$, $j \in J_0(\bar{x}, \bar{y})$ have the same sign, we get from (22)

$$sign(\Delta_{ij}) = -sign(\Delta_{ji}), \quad i, j = 1, \ldots, p.$$

Hence,

$$\delta_j > 0 \quad \text{for all } j \in \{1, \ldots, p\} \backslash \{q\}.$$

Finally, in this case we get locally around $(\bar{x}, \bar{y})$:

$$\Sigma_{min} = \left\{ (x, y^q(x)) \mid x \geq \bar{x} \text{ (resp. } x \leq \bar{x}) \text{ if } \gamma_q = +1 \text{ (resp. } \gamma_q = -1), \quad \delta_q = 0 \right\}.$$

We refer to this case as Type 5-1.

*Case 2: $\mu_j$, $j \in J_0(\bar{x}, \bar{y})$ have different signs*

The separation argument implies MFCQ to be satisfied at $(\bar{x}, \bar{y})$. Hence, a local minimizer $\bar{y}$ for $L(\bar{x})$ is also a KKT-point and $(\bar{x}, \bar{y}) \in \Sigma_{\text{KKT}}$. Due to MFCQ, $(\bar{x}, \bar{y})$ is not a boundary point of $\Sigma_{\text{KKT}}$. Thus, there exist $q, r \in \{1, \ldots, p\}$, $q \neq r$ such that

$$\delta_q = 0, \quad \gamma_q = -1 \quad \text{and} \quad \delta_r = 0, \quad \gamma_r = +1.$$

Moreover, such $q, r$ are unique due to (22), D4 and definition of $\gamma_j$'s.

In this case we get locally around $(\bar{x}, \bar{y})$:

$$\Sigma_{min} = \left\{ (x, y(x)) \mid y(x) := \left\{ \begin{array}{l} y^q(x), \ x \leq \bar{x} \ (\text{if } \delta_q = 0, \quad \gamma_q = -1) \\ y^r(x), \ x \geq \bar{x} \ (\text{if } \delta_r = 0, \quad \gamma_r = 1) \end{array} \right. \right\}.$$

We refer to the case as Type 5-2.

## 4 Main results

First, we define simplicity of a bilevel programming problem at a feasible point. Recall again that $dim(x) = 1$.

**Definition 4.1** (*Simplicity of bilevel problems*) A bilevel programming problem $U$ (with $dim(x) = 1$) is called simple at $(\bar{x}, \bar{y}) \in M$ if one of the following cases occurs:
Case I: Argmin $L(\bar{x}) = \{\bar{y}\}$ and $(\bar{x}, \bar{y})$ is of Type 1, 2, 4, 5-1 or 5-2,
Case II: Argmin $L(\bar{x}) = \{\bar{y}_1, \bar{y}_2\}$ and $(\bar{x}, \bar{y}_1)$, $(\bar{x}, \bar{y}_2)$ are both of Type 1, additionally it holds:

$$\alpha := sign \left[ \left. \frac{d\,[g(x, y_2(x)) - g(x, y_1(x))]}{dx} \right|_{x=\bar{x}} \right] \neq 0, \tag{23}$$

where $y_1(x)$, $y_2(x)$ are unique local minimizers for $L(x)$ in a neighborhood of $\bar{x}$ with $y_1(\bar{x}) = \bar{y}_1$, $y_2(\bar{x}) = \bar{y}_2$ according to Type 1.

In order to avoid asymptotic effects, let $\mathcal{O}$ denote the set of $(g, h_j, j \in J) \in C^3(\mathbb{R}^{1+m}) \times \left[C^3(\mathbb{R}^{1+m})\right]^{|J|}$ such that

$$B_{g,h}(\bar{x}, c) \text{ is compact for all } (\bar{x}, c) \in \mathbb{R} \times \mathbb{R}, \tag{24}$$

where

$$B_{g,h}(\bar{x}, c) := \{(x, y) \mid \|x - \bar{x}\| \le 1, \quad g(x, y) \le c, \quad y \in M(x)\}.$$

Note that $\mathcal{O}$ is $C^3_s$-open.

Now, we state our main result.

**Theorem 4.1** (Simplicity is generic and stable) *Let $\mathcal{F}$ denote the set of defining functions $(f, g, h_j, j \in J) \in C^3(\mathbb{R}^{1+m}) \times \mathcal{O}$ such that the corresponding bilevel programming problem $U$ is simple at all its feasible points $(\bar{x}, \bar{y}) \in M$. Then, $\mathcal{F}$ is $C^3_s$-open and $C^3_s$-dense in $C^3(\mathbb{R}^{1+m}) \times \mathcal{O}$.*

*Proof* It is well-known from the one-dimensional parametric optimization ([13]) that generically the points of $\Sigma$ are of Types 1-5 as defined above. Moreover, for the points of $M \subset \Sigma$ only Types 1, 2, 4, 5-1 or 5-2 may occur generically (cf. Sect. 3). Further, the appearance of two different $y, z \in$ Argmin $L(x)$ causes one loss of freedom to the equation

$$g(x, y) = g(x, z).$$

From the standard argument by counting the dimension and codimension of the corresponding manifold in multi-jet-space and by applying the Multi-Jet-Transversality Theorem (cf. [14]), we get generically:

$$|\text{Argmin } L(x)| \le 2.$$

Now, $|\text{Argmin } L(x)| = 1$ corresponds to Case I in Definition 4.1. For the case $|\text{Argmin } L(x)| = 2$, we obtain the points of Type 1. It comes from the fact that the appearance of Types 2, 4, 5-1 or 5-2 would cause another loss of freedom due to their degeneracy. Analogously, (23) in Case II is generically valid.

The proof of the openness-part is standard (cf. [14]). □

Using the description of $\Sigma_{min}$ from Sect. 3, a reducible bilevel programming problem $U$ can be locally reduced as follows.

**Theorem 4.2** (Bilevel feasible set and Reduced Problem) *Let the bilevel programming problem $U$ (with $dim(x) = 1$) be simple at $(\bar{x}, \bar{y}) \in M$. Then, locally around $(\bar{x}, \bar{y})$, $U$ is equivalent to the following reduced optimization problem:*

$$\text{Reduced} - \text{Problem}: \quad \underset{(x,y)\in\mathbb{R}^1\times\mathbb{R}^m}{\text{minimize}} \; f(x, y) \quad \text{s.t.} \quad (x, y) \in M_{loc}, \tag{25}$$

*where $M_{loc}$ is given according to the cases in Definition 4.1: Case I, Type 1:*

$$M_{loc} = \{(x, y(x)) \mid x \text{ sufficiently close to } \bar{x}\},$$

*Case I, Type 2:*

$$M_{loc} = \left\{(x, y(x)) \mid y(x) := \left\{ \begin{array}{l} \widetilde{y}(x),\ x \le \bar{x} \\ \widehat{y}(x),\ \bar{x} \le x \end{array} \right\} \text{if } sign(\gamma) = -1\right.$$

or

$$M_{loc} = \left\{(x, y(x)) \mid y(x) := \left\{ \begin{array}{l} \widehat{y}(x),\ x \le \bar{x} \\ \widetilde{y}(x),\ \bar{x} \le x \end{array} \right\} \text{if } sign(\gamma) = +1,\right.$$

*Case I, Type 4:*

$$M_{loc} = \{(x(\lambda), y(\lambda)) \mid \lambda \le 0\},$$

*Case I, Type 5-1:*

$$\begin{aligned} M_{loc} = \big\{&(x, y^q(x)) \mid x \ge \bar{x} \ (\text{resp.} x \le \bar{x}) \textit{ if } \gamma_q \\ &= -1 \ (\text{resp.} \gamma_q = +1), \quad \delta_q = 0\big\}, \end{aligned}$$

*Case I, Type 5-2:*

$$M_{loc} = \left\{(x, y(x)) \mid y(x) := \left\{ \begin{array}{l} y^q(x),\ x \le \bar{x} \ (\text{if} \delta_q = 0, \quad \gamma_q = -1) \\ y^r(x),\ x \ge \bar{x} \ (\text{if} \delta_r = 0, \quad \gamma_r = 1) \end{array} \right._s \right\},$$

*Case II:*

$$M_{loc} = \{(x, y_1(x)) \mid x \ge \bar{x} \ (\text{resp. } x \le \bar{x}) \text{ if } \alpha = +1 \ (\text{resp. } \alpha = -1)\}.$$

*We refer to Sect. 3 for details on Types 1, 2, 4. 5-1 and 5-2.* □

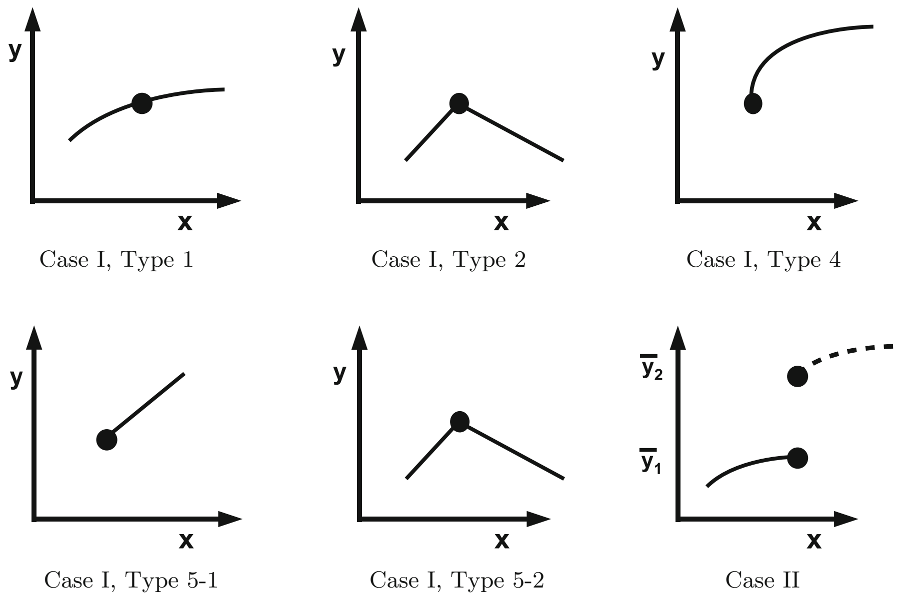



**Fig. 1** Bilevel feasible set $M_{loc}$ from Theorem 4.2

In each case one of the possibilities for $M_{loc}$ is depicted in Fig. 1.

Theorem 4.2 allows to deduce optimality criteria for a reducible bilevel programming problem. In fact, the set $M_{loc}$ from *Reduced-Problem* is the feasible set of either a standard nonlinear optimization problem—NLP—(Cases I, Type 1, 4, 5-1 and Case II) or a mathematical programming problem with complementarity constraints—MPCC—(Cases I, Type 2 and 5-2). Hence, we only need to use the corresponding optimality concepts of a Karush–Kuhn–Tucker point (for NLP) and of a S-stationary point (for MPCC), cf. [15] for the latter concept.

**Theorem 4.3** (First-order optimality for simple bilevel problem) *Let a bilevel programming problem $U$ (with $dim(x) = 1$) be simple at its local minimizer $(\bar{x}, \bar{y}) \in M$. Then, according to the cases in Theorem 4.2 we obtain: Case I, Type 1:*

$$D_x f(\bar{x}, \bar{y}) + D_y f(\bar{x}, \bar{y}) \cdot D_x y(\bar{x}) = 0,$$

*Case I, Type 2:*

$$\begin{aligned} \left[D_x f(\bar{x}, \bar{y}) + D_y f(\bar{x}, \bar{y}) \cdot D_x \widetilde{y}(\bar{x})\right] &\leq 0, \\ \left[D_x f(\bar{x}, \bar{y}) + D_y f(\bar{x}, \bar{y}) \cdot D_x \widehat{y}(\bar{x})\right] &\leq 0. \end{aligned}$$

*Case I, Type 4:*

$$D_x f(\bar{x}, \bar{y}) \cdot D_\lambda x(0) + D_y f(\bar{x}, \bar{y}) \cdot D_\lambda y(0) \leq 0,$$

*Case I, Type 5-1:*

$$D_x f(\bar{x}, \bar{y}) + D_y f(\bar{x}, \bar{y}) \cdot D_x y^q(\bar{x}) \geq 0, \text{if} \gamma_q = -1, \quad \delta_q = 0$$

*or*

$$D_x f(\bar{x}, \bar{y}) + D_y f(\bar{x}, \bar{y}) \cdot D_x y^q(\bar{x}) \leq 0, \text{if} \gamma_q = +1, \quad \delta_q = 0$$

*Case I, Type 5-2:*

$$\begin{aligned} \left[D_x f(\bar{x}, \bar{y}) + D_y f(\bar{x}, \bar{y}) \cdot D_x y^q(\bar{x})\right] \leq 0, \\ \left[D_x f(\bar{x}, \bar{y}) + D_y f(\bar{x}, \bar{y}) \cdot D_x y^r(\bar{x})\right] \leq 0, \end{aligned}$$

*Case II:*

$$D_x f(\bar{x}, \bar{y}) + D_y f(\bar{x}, \bar{y}) \cdot D_x y(\bar{x}) \geq 0, \text{if} \alpha = -1,$$

*or*

$$D_x f(\bar{x}, \bar{y}) + D_y f(\bar{x}, \bar{y}) \cdot D_x y(\bar{x}) \leq 0, \text{if} \alpha = +1$$

*Note that the derivatives of implicit functions above can be obtained from the defining equations as discussed in Sect. 3.* □

The following remarks are launched by anonymous referees.

*Remark 4.1* The simplicity of bilevel problems (cf. Definition 4.1) means, in particular, that if the lower level problem does not have a unique optimal solution for the fixed upper level variable, then the optimal solutions of the lower level problem are of Type 1. Hence, the question arise, whether it is possible to derive generic optimality conditions for the bilevel optimization problems using the approach of implicitly defined lower level solution functions (as done under additional convexity assumptions in [5]). Unfortunately, the answer is negative in general. Indeed, as we see from Cases I, Type 4 and Type 5-1 the optimal solution of the lower level problem can not be parametrized by $x$ in the whole neighborhood of $\bar{x}$ (cf. Fig. 1). It is due to the fact that the solution of the lower level problem for the fixed parameter $\bar{x}$ is not strongly stable (cf. Examples 2.2, 2.3). We point out that the assumption of strong stability at the lower level is rather restrictive, already in case $dim(x) = 1$. □

*Remark 4.2* It is well-known that optimality conditions for bilevel programming can be derived in different ways. Here, we mention only the KKT approach and the value function approach. Due to the KKT approach, the lower level problem is replaced by its Karush–Kuhn–Tucker system (cf. e.g. [8]). In the value function approach a nonsmooth optimal value function appears as a constraint instead of the lower level problem (cf e.g. [17]). Recently a kind of combined approach using both ideas was introduced in [16]. We point out that the comparison of optimality conditions (derived in the literature and from Theorem 4.3) is an interesting and involved issue. The difficulty comes from the fact the optimality conditions from Theorem 4.3 are given according to the classification of possible singularities at the lower level. In other approaches, however, the difficult structure of the lower level problem is tried to be avoided. □

*Remark 4.3* We mention the difference between the reduction of a simple bilevel programming problem as performed in Theorem 4.2 [see (25)] and the notion of constraint reducibility (cf. [3]). The latter is widely used in the nonlinear programming literature and refers to the local description of the constraint

$$G(x) \in K \text{ with } G : \mathbb{R}^n \longrightarrow \mathbb{R}^m \text{ and } K \text{ being a closed convex subset of } \mathbb{R}^m$$

as

$$\mathcal{G}(x) \in \mathcal{C} \text{ with } \mathcal{G} : \mathbb{R}^n \longrightarrow \mathbb{R}^k \text{ and } \mathcal{C} \text{ being a closed convex subset of } \mathbb{R}^k.$$

Such reduction makes sense if e.g. $\mathcal{C}$ becomes a pointed convex closed cone. In the bilevel setting the constraint we reduce is

$$y \in \text{Argmin } L(x).$$

Hence, we rather perform the analysis of singularity types appeared at the lower level than apply smooth coordinate transformations as in the case of constraint reducibility. □

## 5 Towards the case $x \in \mathbb{R}^n$, $n \geq 2$

In the higher dimensional case, i.e. $dim(x) \geq 2$, there will appear more complicated singularities in the description of the feasible set. In particular, we will present stable examples when more than one Lagrange multiplier vanishes. This will be an extension of Type 2 (cf. Examples 5.5, 5.6). On the other hand, we will not be able to describe all generic situations. This obstruction comes from classification in singularity theory. In fact, in one variable $(y)$ there is already a countable infinite list of local minimizers: In the unconstrained case the functions $y^{2k}$, $k \geq 1$ and in the constrained case $y \geq 0$ the functions $y^k$, $k \geq 1$. However, a complete list of local minimizers for functions of two variables or more is even not known. Therefore, we have to bring the upper objective function of the bilevel problem into play as well. If we restrict ourselves to a neighborhood of a (local) solution of the bilevel problem, then the generic situation becomes easier. For example, the above mentioned singularities $y^{2k}$ $(k \geq 2)$ as well as the constrained singularities $y^k$ $(k \geq 3)$, $y \geq 0$, can generically be avoided at local solutions of the bilevel problem. The key idea is explained below in Remark 5.6 and illustrated in Examples 5.7, 5.8.

*Remark 5.4* We note that all singularities appearing for lower dimensional $x$ may reappear at higher dimensional $x$ in a kind of product structure. In fact, the lower dimensional singularity may appear as normal section in the corresponding normal-tangential stratification (cf. [9]). For example, let $x = (x_1, x_2, \ldots, x_n)$ and let the lower level problem $L(x)$ be:

$$L(x): \quad \min_y \, (y - x_1)^2 \quad \text{s.t.} \quad y \geq 0$$

Then, the feasible set $M$ becomes:

$$M = \left\{(x_1, x_2, \ldots, x_n, \max\{x_1, 0\} \,|\, x \in \mathbb{R}^n\right\}$$

and in this particular case we see that $M$ is diffeomorphic to product $\big\{(x_1, \max\{x_1, 0\}) \,|\, x_1 \in \mathbb{R}\big\} \times \mathbb{R}^{n-1}$. □

At this point we come to typical examples with several vanishing Lagrange multipliers. Here, we assume that LICQ at the lower level is fulfilled, that the dimensions of the variables $x$ and $y$ coincide (i.e. $n = m$), that $J_0(\bar{x}, \bar{y}) = m$ and that $\bar{x} = \bar{y} = 0$. Taking the constraints $h_j$ as new coordinates, we may assume that the lower level feasible set $M(0)$ is just the nonnegative orthant. In this setting, the Lagrange multipliers of the lower level function $g$ at the origin just become the partial derivatives with respect to the coordinates $y_j$, $j = 1, \ldots, m$. Now we suppose that all these partial derivatives vanish (generalization of Type 2). Then, the Hessian $D^2_{yy}g(0, 0)$ comes into play and we assume that it is nonsingular. In order that the origin is a (local) minimizer for $L(0)$, a stable condition becomes that the positive cone of the Hessian $D^2_{yy}g(0, 0)$ contains the nonnegative orthant with deleted origin. This gives rise to several combinatorial possibilities, depending on the number of negative eigenvalues of $D^2_{yy}g(0, 0)$. In the next two examples, we restrict ourselves to two dimensions, i.e. $n = m = 2$.

*Example 5.5* In this example the Hessian $D^2_{yy}g(0, 0)$ has two (typically distinct) positive eigenvalues. In particular, $D^2_{yy}g(0, 0)$ is positive definite:

$$\begin{aligned}
&f(x_1, x_2, y_1, y_2) = (-x_1 + 2y_1) + (-x_2 + 2y_2)\\
&L(x_1, x_2): \quad \min_y \, g(x_1, x_2, y_1, y_2) := (y_1 - x_1)^2 + (y_1 - x_1) \cdot (y_2 - x_2) + (y_2 - x_2)^2\\
&\text{s.t.} \quad y_1 \geq 0, \quad y_2 \geq 0.
\end{aligned}$$

In order to obtain the feasible set $M$, we have to consider critical points of $L(x_1, x_2)$ for the following four cases I–IV. These cases result from the natural stratification of the nonnegative orthant in $y$-space:

$$\begin{aligned}
\text{I}: y_1 > 0, y_2 > 0 \quad & \text{II}: y_1 = 0, y_2 > 0\\
\text{III}: y_1 > 0, y_2 = 0 \quad & \text{IV}: y_1 = 0, y_2 = 0.
\end{aligned}$$

It turns out that the feasible set $M$ is piecewise smooth two-dimensional manifold. Moreover, it can be parametrized via the $x$-variable by means of a subdivision of the $x$-space into four regions according to the above cases I–IV, see Fig. 2.

**Fig. 2** Illustration of Example 5.5

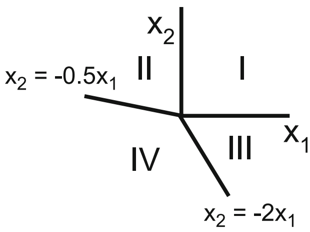
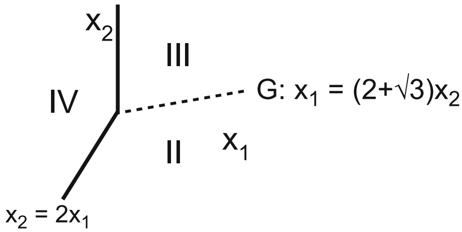



**Fig. 3** Illustration of Example 5.6




On the regions I–IV the corresponding global minimizer $(y_1(\cdot), y_2(\cdot))$ is given by:

$$(y_1(x), y_2(x)) = \begin{cases} (x_1, x_2), & \text{if } (x_1, x_2) \in \text{I}, \\ (0, \frac{x_1}{2} + x_2), & \text{if } (x_1, x_2) \in \text{II}, \\ (\frac{x_2}{2} + x_1, 0), & \text{if } (x_1, x_2) \in \text{III}, \\ (0, 0), & \text{if } (x_1, x_2) \in \text{IV}. \end{cases} \tag{26}$$

In particular, we obtain $M = \{(x, y(x)) \mid y(x) \text{ as in (26)}\}$. A few calculations show that the origin $(0, 0)$ solves the corresponding bilevel problem $U$. □

*Example 5.6* In this example the Hessian $D^2_{yy}g(0, 0)$ has one positive and one negative eigenvalue:

$$f(x_1, x_2, y_1, y_2) = -3x_1 + x_2 + 4y_1 + 5y_2$$

$$L(x_1, x_2): \quad \min_y g(x_1, x_2, y_1, y_2) := (y_1 - x_1)^2 + 4(y_1 - x_1) \cdot y_2 + 3\left(y_2 + \frac{1}{3}x_2\right)^2$$

$$\text{s.t.} \quad y_1 \geq 0, \quad y_2 \geq 0.$$

It is easy to see that $(y_1, y_2) = (0, 0)$ is the global minimizer for $L(0, 0)$. Analogously to Example 5.5 we subdivide the parameter space $(x_1, x_2)$ into regions on which the global minimizer $(y_1(x), y_2(x))$ for L(x) is a smooth function. Here, we obtain three regions II–IV, see Fig. 3. Note that the region corresponding to the case I is empty.

In addition, for the parameters $(x_1, x_2)$ lying on the halfline

$$G : x_1 = (2 + \sqrt{3})x_2, \quad x_1 \geq 0$$

the problem $L(x)$ exhibits two different global minimizers. It is due to the fact that $(y_1, y_2) = (0, 0)$ is a saddlepoint of the objective function $g(0, y_1, y_2)$. Moreover, $(y_1, y_2) = (0, 0)$ is not strongly stable for $L(0, 0)$.

On the regions II–IV and on $G$ the corresponding global minimizers $(y_1(\cdot), y_2(\cdot))$ are given by:

$$(y_1(x), y_2(x)) = \begin{cases} (0, \frac{2}{3}x_1 - \frac{1}{3}x_2), & \text{if } (x_1, x_2) \in \text{II}, \\ (x_1, 0), & \text{if } (x_1, x_2) \in \text{III}, \\ (0, 0), & \text{if } (x_1, x_2) \in \text{IV}, \\ \left\{(0, \frac{2}{3}x_1 - \frac{1}{3}x_2), (x_1, 0)\right\} & \text{if } (x_1, x_2) \in G. \end{cases} \tag{27}$$

Here, $M = \{(x, y(x)) \mid y(x) \text{ as in } (27)\}$. We point out that the bilevel feasible set $M$ is now a two-dimensional nonsmooth Lipschitz manifold with boundary, but it cannot be parametrized by the variable $x$. Again, one calculates that origin (0,0) solves the corresponding bilevel problem $U$. □

*Remark 5.5* Let us consider in Examples 5.5, 5.6 a smooth curve around the origin which traverses the partition of $x$-space in a transversal way, for example a circle $C$. Then, restricted to $C$, the dimension of $x$ reduces to one and we rediscover a simple bilevel problem. □

*Remark 5.6* In order to avoid certain higher order singularities in the description of the feasible set $M$, we have to focus on a neighborhood of (local) solutions of the bilevel problem. The key idea is as follows. Suppose that the feasible set $M$ contains a smooth curve, say $C$, through the point $(\bar{x}, \bar{y}) \in M$. Let the point $(\bar{x}, \bar{y})$ be a local solution of the bilevel problem $U$, i.e. $(\bar{x}, \bar{y})$ is a local minimizer for the objective function $f$ on the set $M$. Then, $(\bar{x}, \bar{y})$ is also a local minimizer for $f$ restricted to the curve $C$. If, in addition, $(\bar{x}, \bar{y})$ is a nondegenerate local minimizer for $f_{|C}$, then we may shift this local minimizer along $C$ by means of a linear perturbation of $f$. After that perturbation with resulting $\widetilde{f}$, the point $(\bar{x}, \bar{y})$ is not any more a local minimizer for $\widetilde{f}_{|C}$ and, hence, it is not any more a local minimizer for $\widetilde{f}_{|M}$. Now, if the singularities in $M$ outside of the point $(\bar{x}, \bar{y})$ are of lower order, then in this way we are able to move away from the higher order singularity. This simple idea was used in particular in [7]. The key point however is to find a smooth curve through a given point of the feasible set $M$. An illustration will be presented in Examples 5.7 and 5.8. In contrast, note that in Examples 5.5 and 5.6 such a smooth curve through the origin $(0, 0)$ does not exist. □

*Example 5.7* Consider the one dimensional functions $y^{2k}$, $k = 1, 2, \ldots$ The origin $y = 0$ is always the global minimizer. For $k = 1$ the latter is nondegenerate (Type 1), but for $k \geq 2$ it is degenerate. Let $k \geq 2$ and $x = (x_1, x_2, \ldots, x_{2k-2})$. Then the function $g(x, y)$, with x as parameter,

$$g(x, y) = y^{2k} + x_{2k-2}y^{2k-2} + x_{2k-3}y^{2k-3} + \cdots + x_1 y$$

is a so-called universal unfolding of the singularity $y^{2k}$. Moreover, the singularities with respect to $y$ have a lower codimension (i.e. lower order) outside the origin $x = 0$ (cf. [1,4]). Consider the unconstrained lower level problem

$$L(x) : \quad \min_y g(x, y)$$

with corresponding bilevel feasible set $M$. Let the smooth curve $C$ in $(x, y)$-space be defined by the equations:

$$x_1 = x_2 = \cdots = x_{2k-3} = 0, \quad ky^2 + (k-1)x_{2k-2} = 0.$$

It is not difficult to see that, indeed, $C$ contains the origin and belongs to the bilevel feasible set $M$. □

*Example 5.8* Consider the one dimensional functions $y^k$, $k \geq 1$ under the constraint $y \geq 0$. The origin $y = 0$ is always the global minimizer. The case $k = 1$ is non-degenerate (Type 1), whereas the case $k = 2$ corresponds to Type 2. Let $k \geq 3$ and $x = (x_1, x_2, \ldots, x_{k-1})$. Then, analogously to Example 5.7, the function $g(x, y)$,

$$g(x, y) = y^k + x_{k-1}y^{k-1} + x_{k-2}y^{k-2} + \cdots + x_1 y, \quad y \geq 0,$$

is the universal unfolding of the (constrained) singularity $y^k$, $y \geq 0$. Consider the constrained lower level problem

$$L(x): \quad \min_y g(x, y) \quad \text{s.t.} \quad y \geq 0$$

with corresponding bilevel feasible set $M$.

In order to find a smooth curve $C$ through the origin and belonging to $M$, we put

$$x_1 = x_2 = \cdots = x_{k-3} = 0.$$

So, we are left with the reduced lower level problem function

$$\widetilde{L}(x_{k-2}, x_{k-1}): \quad \min_y \widetilde{g}(x_{k-2}, x_{k-1}, y) \quad \text{s. t.} \quad y \geq 0$$

with reduced feasible set $\widetilde{M}$, where

$$\widetilde{g}(x_{k-2}, x_{k-1}, y) = y^k + x_{k-1}y^{k-1} + x_{k-2}y^{k-2}.$$

Firstly, let $x_{k-1} < 0$ and $x_{k-2} > 0$ and consider the curve defined by the equation

$$x_{k-2} - \frac{1}{4}x_{k-1}^2 = 0.$$

One calculates, that for points on this curve, the lower level $\widetilde{L}$ has two different global minimizers on the set $y \geq 0$ (with $\widetilde{g}$-value zero), one of them being $y = 0$. Secondly, we note that the set $\{(x_{k-1}, x_{k-2}, 0) \mid x_{k-1} \geq 0, \quad x_{k-2} \geq 0\}$ belongs to $\widetilde{M}$. Altogether, we obtain that the curve $C$ defined by the equations

$$x_1 = x_2 = \cdots = x_{k-3} = y = 0, \quad x_{k-2} - \frac{1}{4}x_{k-1}^2 = 0,$$

belongs to $M$. □

We finally remark that a complete systematic generic description of the feasible bilevel set $M$ in a neighborhood of local solutions of the bilevel problem $U$ for higher $x$-dimensions is a very challenging issue for future research. Another interesting point for future research would be the discovery of a stable generic constraint qualification under which the whole feasible set $M$ might be expected to be a Lipschitz manifold with boundary.

**Acknowledgments** The authors would like to thank the referees for their precise remarks.

## References

1. Arnold, V.I.: Singularity Theory. Cambridge University Press, Cambridge (1981)
2. Bard, J.F.: Practical Bilevel Optimization: Algorithms and Applications. Kluwer, Dordrecht (1998)
3. Bonnans, F., Shapiro, A.: Perturbation Analysis of Optimization Problems. Springer, New York (2000)
4. Bröcker, P., Lander, L.: Differentiable germs and catastrophes. Cambridge University Press, Cambridge (1975)
5. Dempe, S.: A necessary and a sufficient optimality condition for bilevel programming problems. Optimization **25**, 341–354 (1992)
6. Dempe, S.: Foundations of Bilevel Programming. Kluwer, Dordrecht (2002)
7. Dempe, S., Günzel, H., Jongen, H.Th.: On reducibility in bilevel problems. SIAM J. Optim. **20**(2), 718–727 (2009)
8. Dempe, S., Dutta, J.: Is bilevel programming a special case of a mathematical program with complementarity constraints? Math. Program. **131**, 37–48 (2012)
9. Goresky, M., MacPherson, R.: Stratified Morse Theory. Springer, Berlin (1988)
10. Hirsch, M.W.: Differential Topology. Springer, Berlin (1976)
11. Kojima, M.: Strongly stable stationary solutions in nonlinear programs. In: Robinson, S.M. (ed.) Analysis and Computation of Fixed Points, pp. 93–138. Academic Press, New York (1980)
12. Jongen, H.Th.: Parametric optimization: critical points and local minima. Lect. Appl. Math. **26**, 317–335 (1990)
13. Jongen, H.Th., Jonker, P., Twilt, F.: Critical sets in parametric optimization. Math. Program. **34**, 333–353 (1986)
14. Jongen, H.Th., Jonker, P., Twilt, F.: Nonlinear Optimization in Finite Dimensions. Kluwer, Dordrecht (2000)
15. Scheel, H., Scholtes, S.: Mathematical programs with complementarity constraints: stationarity, optimality, and sensitivity. Math. Oper. Res. **25**(1), 1–22 (2000)
16. Ye, J.J., Zhu, D.L.: New necessary optimality conditions for bilevel programs by combined MPEC and the value function approach. SIAM J. Optim. **20**, 1885–1905 (2010)
17. Ye, J.J., Zhu, D.L.: Optimality conditions for bilevel programming problems. Optimization **33**, 9–27 (1995)