



TECHNISCHE UNIVERSITÄT
BERGAKADEMIE FREIBERG

Die Ressourcenuniversität. Seit 1765.

Fakultät für Mathematik und Informatik

Preprint 2014-12

**Stephan Dempe
and Patrick Mehlitz**

A remark on semivectorial bilevel
programming and an application in
semivectorial bilevel optimal control

ISSN 1433-9307

Stephan Dempe and Patrick Mehlitz

A remark on semivectorial bilevel programming
and an application in semivectorial bilevel optimal control

TU Bergakademie Freiberg

Fakultät für Mathematik und Informatik

Prüferstraße 9

09596 FREIBERG

<http://www.mathe.tu-freiberg.de>

ISSN 1433 – 9307

Herausgeber: Dekan der Fakultät für Mathematik und Informatik

Herstellung: Medienzentrum der TU Bergakademie Freiberg

A remark on semivectorial bilevel programming and an application in semivectorial bilevel optimal control

Stephan Dempe · Patrick Mehlitz

Abstract We consider a general optimistic semivectorial bilevel programming problem in Banach spaces and provide a result on its relationship to a corresponding scalarized bilevel programming problem. Meanwhile, a mistake in an article by Dempe et. al is depicted and corrected. Afterwards, we apply the derived theory in order to characterise the local optimal solutions of a special semivectorial bilevel optimal control problem. For that purpose it is transformed into a scalar bilevel optimal control problem which can be tackled with the KKT-approach of bilevel programming. Subsequently, we reveal the relationship of the emerging surrogate problem and the original semivectorial bilevel optimal control problem. Finally, we derive a Pontryagin-type necessary optimality condition for the resulting single-level optimal control problem and justify that it is a necessary optimality condition for the original semivectorial bilevel optimal control problem as well.

Keywords Bilevel Programming · Multiobjective Optimization · Optimization in Banach Spaces · Non-smooth Optimization · Optimal Control · Pontryagin Maximum Principle

1 Introduction and notation

This paper is dedicated to the study of the optimistic semivectorial bilevel programming problem

$$\begin{aligned} F(x, y) &\rightarrow \min_{x, y} \\ G(x) &\in -K_u \\ y &\in \Psi_{we}(x) \end{aligned} \tag{1}$$

where $\Psi_{we}: X \rightarrow \mathfrak{P}(Y)$ denotes the solution-set-mapping of the multiobjective parametric optimization problem

$$\begin{aligned} f(x, y) &\rightarrow \min_y \\ g(x, y) &\in -K_l \end{aligned} \tag{2}$$

in the sense of *weakly efficient* solutions. Such problems were already investigated in the finite-dimensional setting in [10]. Before we start to discuss problem (1) in more detail, we postulate our standing assumptions on that problem and provide an overview of the basic notation used throughout this article.

Forthwith, the following assumptions on (1) shall hold:

- $F: X \times Y \rightarrow \mathbb{R}$, $G: X \rightarrow Z_u$, $f: X \times Y \rightarrow \mathbb{R}^k$ and $g: X \times Y \rightarrow Z_l$ are continuous mappings between Banach spaces X , Y , Z_u , and Z_l (and \mathbb{R} as well as \mathbb{R}^k , obviously),

S. Dempe

Faculty of Mathematics and Computer Science, Technical University Bergakademie Freiberg, Germany
E-mail: dempe@math.tu-freiberg.de

P. Mehlitz (Corresponding author)

Faculty of Mathematics and Computer Science, Technical University Bergakademie Freiberg, Germany
E-mail: mehlitz@math.tu-freiberg.de

- the sets $K_u \subseteq Z_u$ and $K_l \subseteq Z_l$ are ordering cones (i.e. non-empty, closed, convex and pointed cones),
- the mapping f possesses the k component mappings $f_1, \dots, f_k: X \times Y \rightarrow \mathbb{R}$ and for any $x \in X$ the mappings $f_1(x, \cdot), \dots, f_k(x, \cdot)$ are convex,
- for any $x \in X$ the mapping $g(x, \cdot)$ is K_l -convex, i.e.

$$\forall y, y' \in Y \forall \alpha \in [0, 1]: \quad g(x, \alpha \cdot y + (1 - \alpha) \cdot y') - \alpha \cdot g(x, y) - (1 - \alpha) \cdot g(x, y') \in -K_l$$

holds true,

- the minimization in (2) happens with respect to (w.r.t.) the natural ordering cone in \mathbb{R}^k , i.e. the cone of all vectors from \mathbb{R}^k which possess non-negative components.

Our main purpose concerning (1) is to clarify its relationship to a surrogate scalar bilevel programming problem derived from (1) by applying the *weighted-sum-scalarization-technique* to the lower level problem (2). Furthermore, we want to present an application of semivectorial bilevel programming in optimal control.

We organize this paper as follows: In the rest of this first section we provide basic notations, definitions and a result we are going to use. Afterwards, we derive the aforementioned scalarized bilevel programming problem, which corresponds to (1), in Section 2. By means of an example we show that these two bilevel programming problems do not need to coincide locally, which contradicts Proposition 3.2 in [10]. Subsequently, we present a corrected form of that result and discuss some possibilities to reformulate (1) as a single-level scalar optimization problem. Section 3 is dedicated to the analysis of a special semivectorial bilevel optimal control problem by means of a certain class of local minimizers comprising global optimal solutions. We apply the results of the previous section and some results of optimal control in order to find a corresponding single-level optimal control problem. Afterwards, we exploit a general Pontryagin-type optimality condition from [25], which provides a rather weak necessary optimality condition for the original semivectorial bilevel optimal control problem.

Firstly, we are going to introduce several notations and definitions we use throughout this paper. Forthwith, let \mathbb{N} , \mathbb{R} , \mathbb{R} , \mathbb{R}^n , $\mathbb{R}_0^{n,+}$, and $\mathbb{R}^{m \times n}$ denote the natural numbers, the real numbers, the extended real line (i.e. $\mathbb{R} \cup \{-\infty, +\infty\}$), the space of real vectors with n components, the cone of all vectors in \mathbb{R}^n possessing non-negative components, and the set of all real matrices with m rows and n columns, respectively. If not stated otherwise, we equip the linear spaces \mathbb{R} , \mathbb{R}^n and $\mathbb{R}^{m \times n}$ with the infinity-norm, which will be denoted by $\|\cdot\|_\infty$. Let us stipulate that the relation \leq should be interpreted componentwise when comparing objects of same dimension. For an arbitrary matrix $Q \in \mathbb{R}^{m \times n}$, $Q^\top \in \mathbb{R}^{n \times m}$ denotes its transpose. For $i \in \{1, \dots, n\}$ we use $\mathbf{e}_i \in \mathbb{R}^n$ to express the i -th unit vector in \mathbb{R}^n . Furthermore, we define the n -dimensional all-ones vector $\mathbf{e} \in \mathbb{R}^n$ by $\mathbf{e} := \sum_{i=1}^n \mathbf{e}_i$.

Let X, Y and Z be real Banach spaces. We represent the norm defined in X by $\|\cdot\|_X$ while \mathbf{o}_X represents the zero vector in this space. Additionally, we use \mathbb{U}_X to denote its open unit ball and define $\mathbb{U}_X^\varepsilon(\hat{x}) := \{\hat{x}\} + \varepsilon \cdot \mathbb{U}_X$ for any choice of $\hat{x} \in X$ and $\varepsilon > 0$. Let X^* be the dual space of X . We define the corresponding dual pairing by $\langle \cdot, \cdot \rangle: X \times X^* \rightarrow \mathbb{R}$. For any set $A \subseteq X$ we identify its convex hull, conic hull and its power set with $\text{conv}(A)$, $\text{cone}(A)$ and $\mathfrak{P}(A)$, respectively. If A is non-empty, we denote its dual cone by $A^D := \{x^* \in X^* \mid \forall x \in A: \langle x, x^* \rangle \geq 0\}$. Assume that A is non-empty as well as convex and choose $a \in A$ arbitrarily. Then the normal cone to A at a is given by $\mathcal{N}_A(a) := -(A - \{a\})^D$. Let $B \subseteq X^*$ be a non-empty set. Its weak* closure (i.e. its closure w.r.t. the weak* convergence in X^*) is denoted by $\text{cl}^*(B)$ (cf. [26]). We equip the product space $X \times Y$ of X and Y with the maximum norm, i.e. $\|(x, y)\|_{X \times Y} := \max\{\|x\|_X; \|y\|_Y\}$ holds for all points $(x, y) \in X \times Y$. By X^n we denote the product space $\prod_{\nu=1}^n X$. The set $L(X, Y)$ comprises all bounded linear operators which map from X to Y . Let $\Theta: X \rightarrow \mathfrak{P}(Y)$ be a set-valued mapping. Then $\text{dom}(\Theta) := \{x \in X \mid \Theta(x) \neq \emptyset\}$ and $\text{graph}(\Theta) := \{(x, y) \in X \times Y \mid y \in \Theta(x)\}$ denote the domain and the graph of Θ , respectively. We call Θ closed at a certain point $\hat{x} \in X$ provided for any sequences $(x_\nu) \subseteq X$ converging to \hat{x} and $(y_\nu) \subseteq Y$ converging to some point $\hat{y} \in Y$ such that $y_\nu \in \Theta(x_\nu)$ holds true for all $\nu \in \mathbb{N}$ we have $\hat{y} \in \Theta(\hat{x})$. Furthermore, we refer to Θ as a closed mapping if it is closed at any point $x \in X$. Moreover, Θ is said to be lower semicontinuous at $\hat{x} \in \text{dom}(\Theta)$ provided for any open set $\Omega \subseteq Y$ satisfying $\Theta(\hat{x}) \cap \Omega \neq \emptyset$ we find $\varepsilon > 0$ such that $\Theta(x) \cap \Omega \neq \emptyset$ holds true for all $x \in \mathbb{U}_X^\varepsilon(\hat{x})$ (cf. [2] and [3] for a detailed introduction to the topic of set-valued mappings).

Let $I \subseteq \mathbb{R}$ be an arbitrary interval. Furthermore, choose mappings $\alpha: I \rightarrow \mathbb{R}^n$, $\beta: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^k$, and $\gamma: X \times Y \rightarrow Z$ which are weakly differentiable, twice continuously differentiable, and continuously Fréchet differentiable, respectively. Then $\dot{\alpha}$ represents the weak derivative of α while for any

$(\hat{v}, \hat{w}) \in \mathbb{R}^n \times \mathbb{R}^m$ we denote by $\nabla\beta(\hat{v}, \hat{w}) \in \mathbb{R}^{k \times (n+m)}$ and $\nabla_v\beta(\hat{v}, \hat{w}) \in \mathbb{R}^{k \times n}$ the Jacobian and partial Jacobian w.r.t. v of β at (\hat{v}, \hat{w}) . Furthermore, in the case $k = 1$ we write $\nabla^2\beta(\hat{v}, \hat{w}) \in \mathbb{R}^{(n+m) \times (n+m)}$, $\nabla_{vv}^2\beta(\hat{v}, \hat{w}) := \nabla_v(\nabla_v\beta)(\hat{v}, \hat{w}) \in \mathbb{R}^{n \times n}$, and $\nabla_{vw}^2\beta(\hat{v}, \hat{w}) := \nabla_w(\nabla_v\beta)(\hat{v}, \hat{w}) \in \mathbb{R}^{n \times m}$ in order to express the Hessian, the Hessian w.r.t. v , and the mixed Hessian w.r.t. v and w of β at (\hat{v}, \hat{w}) , respectively. For an arbitrary point $(\hat{x}, \hat{y}) \in X \times Y$ we denote by $\gamma'(\hat{x}, \hat{y}) \in L(X \times Y, Z)$ and $\gamma'_x(\hat{x}, \hat{y}) \in L(X, Z)$ the Fréchet derivative and the partial Fréchet derivative w.r.t. x of γ at (\hat{x}, \hat{y}) , respectively. If $\delta: X \rightarrow Y$ and $\eta: Y \rightarrow Z$ are arbitrary mappings, we express their composition by $\delta \circ \eta: X \rightarrow Z$. Finally, for any non-empty set $\Omega \subseteq X$ we stipulate $\delta(\Omega) := \{\delta(x) \mid x \in \Omega\}$.

Let $[0, T]$ be a non-empty real interval. Then $L_1[0, T]$ denotes the Banach space of all scalar functions on $[0, T]$ which are measurable in the sense of Lebesgue. We define a norm in that space by means of $\|x\|_{L_1[0, T]} := \int_0^T |x(s)| ds$ for arbitrary $x \in L_1[0, T]$. The space $C[0, T]$ contains all continuous scalar functions defined on $[0, T]$. This space is equipped with the maximum-norm. The Sobolev space $W_{1,1}[0, T]$ comprises all scalar absolutely continuous functions on $[0, T]$ possessing a weak derivative which belongs to $L_1[0, T]$. Then $\|x\|_{W_{1,1}[0, T]} := \|x\|_{L_1[0, T]} + \|\dot{x}\|_{L_1[0, T]}$ for arbitrary $x \in W_{1,1}[0, T]$ defines a norm in $W_{1,1}[0, T]$. Finally, we declare $L_1^n[0, T] := (L_1[0, T])^n$, $C^n[0, T] := (C[0, T])^n$, and $W_{1,1}^n[0, T] := (W_{1,1}[0, T])^n$.

Recall the following version of an embedding theorem for Sobolev spaces.

Lemma 1.1 (*Sobolev Embedding Theorem, cf. [1]*)

Let $T > 0$ and a positive integer $n \in \mathbb{N}$ be fixed. Then there exists a constant $c > 0$ such that the following inequality holds true:

$$\forall x \in W_{1,1}^n[0, T]: \quad \|x\|_{C^n[0, T]} \leq c \cdot \|x\|_{W_{1,1}^n[0, T]}.$$

2 Derivation of surrogate problems

We start this section recalling the definition of weak efficiency (cf. [17]).

Definition 2.1 (Weak efficiency)

Let $\vartheta: V \rightarrow \mathbb{R}^k$ be an arbitrary mapping, $\Omega \subseteq V$ be a non-empty subset of a real Banach space V , and $C \subseteq \mathbb{R}^k$ be an ordering cone possessing a non-empty interior. A point $\hat{v} \in V$ is called a **weakly efficient** point w.r.t. the ordering cone C of the multiobjective optimization problem

$$\min\{\vartheta(v) \mid v \in \Omega\} \tag{3}$$

provided the following holds true:

$$\vartheta(\Omega) \cap (\{\vartheta(\hat{v})\} - \text{int}(C)) = \emptyset.$$

The above definition can be simplified if the ordering cone $C := \mathbb{R}_0^{k,+}$ is chosen. Let $\vartheta_1, \dots, \vartheta_k: V \rightarrow \mathbb{R}$ denote the k component functions of ϑ . A point $\hat{v} \in \Omega$ is weakly efficient w.r.t. the ordering cone $\mathbb{R}_0^{k,+}$ (or, for short, weakly efficient) for problem (3) if and only if there is no $v \in \Omega$ such that

$$\forall i \in \{1, \dots, k\}: \quad \vartheta_i(v) < \vartheta_i(\hat{v})$$

holds true (cf. [14]).

Now we draw back our attention to the parametric optimization problem (2). First, we define a set-valued mapping $\Gamma: X \rightarrow \mathfrak{P}(Y)$ which describes the feasible set of this problem:

$$\forall x \in X: \quad \Gamma(x) := \{y \in Y \mid g(x, y) \in -K_l\}.$$

In the following we call Γ the feasible-set-mapping of problem (2). Since $g(x, \cdot)$ is assumed to be K_l -convex for any choice of $x \in X$, the images of Γ are convex sets. Using the above characterisation of weak efficiency (w.r.t. the ordering cone $\mathbb{R}_0^{k,+}$) and the notion of Γ we are able to depict the solution-set-mapping $\Psi_{we}: X \rightarrow \mathfrak{P}(Y)$ of (2) in more detail:

$$\forall x \in X: \quad \Psi_{we}(x) = \{y \in \Gamma(x) \mid \nexists \tilde{y} \in \Gamma(x) \forall i \in \{1, \dots, k\}: f_i(x, \tilde{y}) < f_i(x, y)\}.$$

Although the above notion of Ψ_{we} provides a proper geometrical view on that set-valued mapping, we have to think of an efficient way to compute its image sets. This problem can be dealt with using different scalarization approaches (cf. [14] and [17]). Since problem (2) is a convex optimization problem for fixed parameter x , the weighted-sum-scalarization seems to be an appropriate scalarization technique to tackle problem (2).

Let us introduce a set $\mathbb{S}^{k,+} \subseteq \mathbb{R}^k$ of scalarization parameters by:

$$\mathbb{S}^{k,+} := \left\{ z \in \mathbb{R}_0^{k,+} \mid z^\top \mathbf{e} = 1 \right\}.$$

Recall that $\mathbf{e} \in \mathbb{R}^k$ denotes the all-ones vector. Now for any $z \in \mathbb{S}^{k,+}$ we define a scalarized version of problem (2) as follows:

$$\begin{aligned} z^\top f(x, y) &\rightarrow \min_y \\ g(x, y) &\in -K_l. \end{aligned} \tag{4}$$

Since the above problem is a scalar optimization problem now, we are able to define its optimal-value-function $\varphi: X \times \mathbb{S}^{k,+} \rightarrow \overline{\mathbb{R}}$ and its solution-set-mapping $\Psi: X \times \mathbb{S}^{k,+} \rightarrow \mathfrak{P}(Y)$ as given below:

$$\begin{aligned} \forall x \in X \forall z \in \mathbb{S}^{k,+}: \quad \varphi(x, z) &:= \inf_y \{ z^\top f(x, y) \mid y \in \Gamma(x) \} \\ \Psi(x, z) &:= \{ y \in \Gamma(x) \mid z^\top f(x, y) \leq \varphi(x, z) \}. \end{aligned} \tag{5}$$

Note that for any choice of $x \in X$ and $z \in \mathbb{S}^{k,+}$ (4) is a convex optimization problem and hence, the mapping Ψ possesses convex images as well. In fact, this does not necessarily hold true if $z \in \mathbb{R}^k$ is arbitrarily chosen. It is well-known from multicriteria optimization (cf. [14]) that for any $x \in X$ and $z \in \mathbb{S}^{k,+}$ we have $\Psi(x, z) \subseteq \Psi_{we}(x)$. Moreover, due to the convexity of the set $\{f(x, y) \mid y \in \Gamma(x)\} + \mathbb{R}_0^{k,+}$ for any $x \in X$ we have the following result.

Lemma 2.1 (Theorem 5.13 in [17])

Choose an arbitrary point $x \in X$. Then the following formula holds true:

$$\Psi_{we}(x) = \bigcup_{z \in \mathbb{S}^{k,+}} \Psi(x, z).$$

The above lemma allows us to reformulate problem (1) in the way stated below:

$$\begin{aligned} F(x, y) &\rightarrow \min_{x, y} \\ G(x) &\in -K_u \\ y &\in \bigcup_{z \in \mathbb{S}^{k,+}} \Psi(x, z). \end{aligned} \tag{6}$$

Obviously, this equivalent reformulation of (1) possesses a structure, which is similar to a scalar bilevel optimization problem. Unfortunately, z does not play the role of a variable in (6) which makes it very complex to derive any type of optimality condition for the latter problem. Hence, it seems obvious to eliminate this difficulty replacing (6) by:

$$\begin{aligned} F(x, y) &\rightarrow \min_{x, y, z} \\ G(x) &\in -K_u \\ -z &\in -\mathbb{R}_0^{k,+} \\ z^\top \mathbf{e} &= 1 \\ y &\in \Psi(x, z). \end{aligned} \tag{7}$$

Observe that (7) is an optimistic scalar bilevel programming problem. This reformulation of problem (1) is studied for instance in [6, 10] or [19]. Especially in the article [10] the authors provide a result which states the equivalence of the bilevel programming problems (1) and (7) by means of local and global optimal solutions provided the mapping Ψ is closed and all the considered Banach spaces are finite-dimensional. Unfortunately, the fact that this result does not hold in general can be easily demonstrated by simple examples.

Example 2.1

Consider the semivectorial bilevel programming problem

$$\begin{aligned} y - x &\rightarrow \min_{x,y} \\ 0 &\leq x \leq 1 \\ y &\in \Psi_{we}(x) \end{aligned}$$

where the lower level problem is stated below:

$$\begin{aligned} \left. \begin{array}{l} xy \\ 1 - y \end{array} \right\} &\rightarrow \min_y \\ 0 &\leq y \leq 1. \end{aligned}$$

Its unique global optimal solution is given by $(\hat{x}, \hat{y}) = (1, 0)$ and there does not exist a local optimal solution different from (\hat{x}, \hat{y}) . However, the corresponding problem (7) possesses the local optimal solution $(\tilde{x}, \tilde{y}, \tilde{z}_1, \tilde{z}_2) = (1, 1, 0, 1)$.

We start by computing the set of weakly efficient points of the lower level problem for any $x \in [0, 1]$. It is easy to see that $\Psi_{we}(x) = [0, 1]$ holds true for any $x \in [0, 1]$ and hence (\hat{x}, \hat{y}) is the unique local and global optimal solution of the above problem.

From the definition of the set $\mathbb{S}^{2,+}$ of scalarization parameters we have $\mathbb{S}^{2,+} = \text{conv}(\{(0, 1), (1, 0)\})$. Hence, for any $z = (z_1, z_2) \in \mathbb{S}^{k,+}$ we can find $s_z \in [0, 1]$ such that $z_1 = s_z$ and $z_2 = 1 - s_z$ is satisfied. Consequently, in order to reduce the number of parameters in the scalarized lower level problem

$$\begin{aligned} z_1 xy + z_2(1 - y) &\rightarrow \min_y \\ 0 &\leq y \leq 1 \end{aligned} \tag{8}$$

we can consider

$$\begin{aligned} s_z xy + (1 - s_z)(1 - y) &\rightarrow \min_y \\ 0 &\leq y \leq 1 \end{aligned} \tag{9}$$

where $(x, s_z) \in [0, 1]^2$ is the parameter. After solving this problem we only have to countermand the transformation of the parameter z .

One may check that the solution-set-mapping $\Psi: \mathbb{R} \times \mathbb{S}^{2,+} \rightarrow \mathfrak{P}(\mathbb{R})$ of (8) is given as follows for any $x^* \in [0, 1]$ and $(s_z^*, 1 - s_z^*) \in \mathbb{S}^{2,+}$:

$$\Psi(x^*, s_z^*, 1 - s_z^*) = \begin{cases} \{0\} & ; (x^*, s_z^*) \in \left\{ (x, s_z) \mid x \in (0, 1], s_z \in \left(\frac{1}{x+1}, 1 \right] \right\} \\ [0, 1] & ; (x^*, s_z^*) \in \left\{ (x, s_z) \mid x \in [0, 1], s_z = \frac{1}{x+1} \right\} \\ \{1\} & ; (x^*, s_z^*) \in \left\{ (x, s_z) \mid x \in [0, 1], s_z \in \left[0, \frac{1}{x+1} \right) \right\}. \end{cases}$$

Now consider the point $(\tilde{x}, \tilde{y}, \tilde{z}_1, \tilde{z}_2)$ and fix $\varepsilon := \frac{1}{4}$. Suppose there exists a point (x, y, z_1, z_2) feasible for the corresponding problem (7), contained in the ε -neighbourhood of $(\tilde{x}, \tilde{y}, \tilde{z}_1, \tilde{z}_2)$, which satisfies $y - x = F(x, y) < 0 = F(\tilde{x}, \tilde{y})$. Then we have $x \in \left(\frac{3}{4}, 1 \right]$ and $y \in \left(\frac{3}{4}, 1 \right)$. From the above representation of Ψ we automatically have $s_z = \frac{1}{x+1}$ and hence:

$$\|(z_1, z_2) - (\tilde{z}_1, \tilde{z}_2)\|_\infty = \max \left\{ \left| \frac{1}{x+1} - 0 \right|; \left| \frac{x}{x+1} - 1 \right| \right\} \geq \frac{1}{2} > \varepsilon.$$

This inequality contradicts the choice of (x, y, z_1, z_2) in an ε -neighbourhood of the reference point $(\tilde{x}, \tilde{y}, \tilde{z}_1, \tilde{z}_2)$. Hence, the latter point is a local optimal solution of the problem

$$\begin{aligned} y - x &\rightarrow \min_{x,y,z} \\ 0 &\leq x \leq 1 \\ -z_1 &\leq 0 \\ -z_2 &\leq 0 \\ z_1 + z_2 &= 1 \\ y &\in \Psi(x, z_1, z_2) \end{aligned} \tag{10}$$

but (\tilde{x}, \tilde{y}) is no local optimal solution of the original semivectorial bilevel programming problem. Note that the mapping Ψ is closed everywhere on $[0, 1] \times \mathbb{S}^{2,+}$. Hence, this example shows that Proposition 3.2 in [10] is not correct in general. \square

In the following we present a theorem which shows that the problems (1) and (7) are always equivalent w.r.t. global optimal solutions. When considering local optimal solutions, one has to postulate a closedness assumption on Ψ and another restrictive property of problem (7). In order to write the following as short as possible, we introduce another set-valued mapping $\Theta: X \times Y \rightarrow \mathfrak{P}(\mathbb{R}^k)$ by:

$$\forall x \in X \forall y \in Y: \quad \Theta(x, y) := \{z \in \mathbb{S}^{k,+} \mid y \in \Psi(x, z)\}.$$

It is easy to see from Lemma 2.1 that $\text{dom}(\Theta) = \text{graph}(\Psi_{we})$ holds true. Moreover, Θ possesses convex and compact image sets.

Theorem 2.1

1. Let (\hat{x}, \hat{y}) be a global (local) optimal solution of (1). Then for any $z \in \Theta(\hat{x}, \hat{y})$ the point (\hat{x}, \hat{y}, z) is a global (local) optimal solution of (7).
2. Let $(\hat{x}, \hat{y}, \hat{z})$ be a global optimal solution of (7). Then (\hat{x}, \hat{y}) is a global optimal solution of (1).
3. Let (\hat{x}, \hat{y}, z) be a local optimal solution of (7) for all $z \in \Theta(\hat{x}, \hat{y})$ and assume that Ψ is closed at all points (\hat{x}, \bar{z}) where $\bar{z} \in \mathbb{S}^{k,+}$ holds true. Then (\hat{x}, \hat{y}) is a local optimal solution of (1).

Proof

1. We only prove this statement for local optimal solutions. One may choose $\varepsilon = +\infty$ for the global case.
Since (\hat{x}, \hat{y}) is a local optimal solution of (1), we can find a scalar $\varepsilon > 0$ such that for all points $(x, y) \in \mathbb{U}_{X \times Y}^\varepsilon(\hat{x}, \hat{y})$ which are feasible for (1) we have $F(x, y) \geq F(\hat{x}, \hat{y})$. Suppose there exists $\hat{z} \in \Theta(\hat{x}, \hat{y})$ such that $(\hat{x}, \hat{y}, \hat{z})$ is not a local optimal solution of (7). Then we can find sequences $(x_\nu) \subseteq X$ converging to \hat{x} , $(y_\nu) \subseteq Y$ converging to \hat{y} , and $(z_\nu) \subseteq \mathbb{R}^k$ converging to \hat{z} such that (x_ν, y_ν, z_ν) is feasible for problem (7) and $F(x_\nu, y_\nu) < F(\hat{x}, \hat{y})$ holds true for any $\nu \in \mathbb{N}$. Applying Lemma 2.1 from $y_\nu \in \Psi(x_\nu, z_\nu)$ we derive $y_\nu \in \Psi_{we}(x_\nu)$ for any $\nu \in \mathbb{N}$. Due to the convergence of the sequences (x_ν) and (y_ν) this contradicts the local optimality of (\hat{x}, \hat{y}) for (1).
2. Suppose that (\hat{x}, \hat{y}) is not a global optimal solution of (1). Then we find a point $(x, y) \in X \times Y$ feasible for (1) which satisfies $F(x, y) < F(\hat{x}, \hat{y})$. From $y \in \Psi_{we}(x)$ we derive the existence of $z \in \mathbb{S}^{k,+}$ such that (x, y, z) is feasible for problem (7) (cf. Lemma 2.1). This contradicts the global optimality of $(\hat{x}, \hat{y}, \hat{z})$ for problem (7).
3. Suppose that (\hat{x}, \hat{y}) is not a local optimal solution of (1). Then we find sequences $(x_\nu) \subseteq X$ converging to \hat{x} and $(y_\nu) \subseteq Y$ converging to \hat{y} such that (x_ν, y_ν) is feasible for problem (1) and $F(x_\nu, y_\nu) < F(\hat{x}, \hat{y})$ is satisfied for any choice of $\nu \in \mathbb{N}$. Once again we apply Lemma 2.1 to derive the existence of a sequence $(z_\nu) \subseteq \mathbb{R}^k$ such that (x_ν, y_ν, z_ν) is feasible to (7) for any $\nu \in \mathbb{N}$. Then we obviously have $(z_\nu) \subseteq \mathbb{S}^{k,+}$ and since the set $\mathbb{S}^{k,+}$ is compact, (z_ν) possesses a subsequence (z_{ν_μ}) converging to some $\hat{z} \in \mathbb{S}^{k,+}$. Note that due to the closedness of Ψ at (\hat{x}, \hat{z}) we have $\hat{y} \in \Psi(\hat{x}, \hat{z})$ and hence, $(\hat{x}, \hat{y}, \hat{z})$ is a feasible point of (7), which is no local optimal solution of this problem. This contradicts the assumptions of the theorem since we obviously have $\hat{z} \in \Theta(\hat{x}, \hat{y})$ by definition. \square

Recall Example 2.1 by means of the above theorem at the point $(\tilde{x}, \tilde{y}) = (1, 1)$. The corresponding set-valued mapping Ψ is closed everywhere on $[0, 1] \times \mathbb{S}^{2,+}$ and $\Theta(\tilde{x}, \tilde{y}) = \text{conv}(\{(0, 1), (\frac{1}{2}, \frac{1}{2})\})$ holds true. Moreover, for any choice of $s_z \in [0, \frac{1}{2})$ the point $(\tilde{x}, \tilde{y}, s_z)$ is a local optimal solution of problem (9), i.e. $(\tilde{x}, \tilde{y}, s_z, 1 - s_z)$ is a local optimal solution of problem (10). One may check that the point $(\tilde{x}, \tilde{y}, \frac{1}{2})$ is no local optimal solution of (9). Consequently, there exists only one scalarization parameter $\bar{z} = (\frac{1}{2}, \frac{1}{2}) \in \Theta(\tilde{x}, \tilde{y})$ such that $(\tilde{x}, \tilde{y}, \bar{z})$ is not a local optimal solution of (10) and hence, the assumptions of Theorem 2.1 are violated for only one possible choice of the scalarization parameter. However, (\tilde{x}, \tilde{y}) is no local optimal solution of the corresponding semivectorial bilevel programming problem.

The following lemma presents a criterion which can be used to ensure the closedness of the mapping Ψ .

Lemma 2.2

Let the mapping g be partially continuously Fréchet differentiable w.r.t. y . Fix $\hat{x} \in \text{dom}(\Gamma)$ and assume that any point $(\hat{x}, y) \in \text{graph}(\Gamma)$ is partially regular w.r.t. y , i.e. we have:

$$g'_y(\hat{x}, y)[Y] + \text{cone}(K_l + \{g(\hat{x}, y)\}) = Z_l.$$

Then φ is upper semicontinuous and Ψ is closed at any point (\hat{x}, z) where $z \in \mathbb{S}^{k,+}$ is chosen arbitrarily.

Proof We start showing the lower semicontinuity of Γ at \hat{x} . Supposing the contrary there exists an open set $\Omega \subseteq Y$ and a sequence $(x_\nu) \subseteq X$ converging to \hat{x} such that $\Gamma(\hat{x}) \cap \Omega$ is non-empty while $\Gamma(x_\nu) \cap \Omega$ is empty for any $\nu \in \mathbb{N}$. Hence, we find $\hat{y} \in \Gamma(\hat{x})$ and a constant $\delta > 0$ such that $\Gamma(x_\nu) \cap \mathbb{U}_Y^\delta(\hat{y})$ is empty for all $\nu \in \mathbb{N}$. As a consequence, we derive:

$$\forall \nu \in \mathbb{N}: \quad \inf_{y_\nu \in \Gamma(x_\nu)} \|y_\nu - \hat{y}\|_Y \geq \delta. \quad (11)$$

Due to the regularity assumption postulated on (\hat{x}, \hat{y}) we may apply Theorem 1 in [23] and the results in [11] to find $\varepsilon > 0$, $\bar{\delta} > 0$ and a constant $c > 0$ such that $\Gamma(x)$ is non-empty for all $x \in \mathbb{U}_X^\varepsilon(\hat{x})$ and for any such x and $y' \in \mathbb{U}_Y^{\bar{\delta}}(\hat{y})$ we have:

$$\inf_{y \in \Gamma(x)} \|y - y'\|_Y \leq c \cdot \inf_{\xi \in K_l} \|g(x, y') + \xi\|_{Z_l}.$$

Let us fix $y' := \hat{y}$ in the above inequality. Then for sufficiently large $\nu \in \mathbb{N}$ we have:

$$\inf_{y_\nu \in \Gamma(x_\nu)} \|y_\nu - \hat{y}\|_Y \leq c \cdot \inf_{\xi \in K_l} \|g(x_\nu, \hat{y}) + \xi\|_{Z_l} \leq c \cdot \|g(x_\nu, \hat{y}) - g(\hat{x}, \hat{y})\|_{Z_l}.$$

Now we make use of the continuity of g to derive:

$$0 \leq \limsup_{\nu \rightarrow \infty} \inf_{y_\nu \in \Gamma(x_\nu)} \|y_\nu - \hat{y}\|_Y \leq c \cdot \limsup_{\nu \rightarrow \infty} \|g(x_\nu, \hat{y}) - g(\hat{x}, \hat{y})\|_{Z_l} = 0.$$

Obviously, the result above contradicts (11) and hence, Γ is lower semicontinuous at \hat{x} . Moreover, the mapping $(x, y, z) \mapsto z^\top f(x, y)$ is continuous since f is assumed to be continuous. Summarizing these two properties of problem (4) we obtain the upper semicontinuity of the optimal-value-function φ at any point (\hat{x}, z) where $z \in \mathbb{S}^{k,+}$ is chosen arbitrarily (cf. Theorem 4.2.2 in [3]). Now it follows from Theorem 4.2.1 in [3] that Ψ is closed at any such point (\hat{x}, z) . \square

Recall that due to [23] the regularity condition postulated in Lemma 2.2, which is also well-known as RCQ (*Robinson-Constraint-Qualification*) or KRZCQ (*Kurcyusz-Robinson-Zowe-Constraint-Qualification*, cf. [5, 11, 18] or [23]), is equivalent to MFCQ (*Mangasarian-Fromovitz-Constraint-Qualification*) w.r.t. y for problem (4) at any $(\hat{x}, y) \in \text{graph}(\Gamma)$ provided the ordering cone K_l possesses a non-empty interior. In the latter case we can also replace the validity of MFCQ at any point $(\hat{x}, y) \in \text{graph}(\Gamma)$ by the existence of $\tilde{y} \in Y$ which satisfies $g(\hat{x}, \tilde{y}) \in -\text{int}(K_l)$. This condition is referred to as *Slater-Constraint-Qualification* and it is applicable since the constraints of (4) are convex w.r.t. y .

It is well-known in the theory of bilevel programming that the scalar bilevel programming problem (7) can be transformed into an equivalent single-level optimization problem using the optimal-value-function of the corresponding lower level problem. The so-called *optimal-value-reformulation* of (7) is given by

$$\begin{aligned} F(x, y) &\rightarrow \min_{x, y, z} \\ G(x) &\in -K_u \\ -z &\in -\mathbb{R}_0^{k,+} \\ z^\top \mathbf{e} &= 1 \\ z^\top f(x, y) - \varphi(x, z) &\in -\mathbb{R}_0^+ \\ g(x, y) &\in -K_l \end{aligned} \quad (12)$$

where φ is the optimal-value-function of the scalarized lower level problem (4) defined in (5). Be aware that the function φ is likely to be non-smooth, even if the mappings f and g are continuously Fréchet differentiable, and therefore, the handling of (12) seems to be difficult. Additionally, common constraint qualifications like KRZCQ or MFCQ (or at least their non-smooth counterparts) fail at any feasible point of this problem. Furthermore, note that due to Theorem 2.1 problem (12) may possess local optimal solutions which do not correspond to local optimal solutions of problem (1). One may omit the last of the above annoying properties of (12) considering the optimal-value-reformulation of problem (6):

$$\begin{aligned} F(x, y) &\rightarrow \min_{x, y} \\ G(x) &\in -K_u \\ \min_{z \in \mathbb{S}^{k,+}} \{z^\top f(x, y) - \varphi(x, z)\} &\in -\mathbb{R}_0^+ \\ g(x, y) &\in -K_l. \end{aligned} \quad (13)$$

Let us define a function $\psi: X \times Y \rightarrow \overline{\mathbb{R}}$ by:

$$\forall x \in X \forall y \in Y: \quad \psi(x, y) := \min_{z \in \mathbb{S}^{k,+}} \{z^\top f(x, y) - \varphi(x, z)\}.$$

Then problem (13) is equivalent to:

$$\min_{x, y} \{F(x, y) \mid G(x) \in -K_u, \psi(x, y) \in -\mathbb{R}_0^+, g(x, y) \in -K_l\}. \quad (14)$$

Note that ψ is nothing else but the optimal-value-function of the following special parametric optimization problem:

$$\begin{aligned} z^\top f(x, y) - \varphi(x, z) &\rightarrow \min_z \\ -z &\in -\mathbb{R}_0^{k,+} \\ z^\top \mathbf{e} &= 1. \end{aligned} \quad (15)$$

As mentioned in [10] it is easy to see that any feasible point of problem (15) satisfies MFCQ. Furthermore, the feasible set of this problem does not depend on the parameters and is convex and compact. Since the mapping $(x, y, z) \mapsto z^\top f(x, y)$ is at least continuous, the qualitative properties of φ should be an indicator for the properties of the function ψ . Be aware that due to the special structure of problem (4) the function $\varphi(x, \cdot)$ is concave for any choice of $x \in X$ (cf. Proposition 3.5 in [15]). Hence, for fixed $(\hat{x}, \hat{y}) \in X \times Y$ problem (15) is a convex optimization problem which satisfies the Slater-Constraint-Qualification meaning that the (non-smooth) KKT-conditions are necessary and sufficient optimality conditions for (15). Observe that these conditions take the following form for arbitrary $z \in \mathbb{S}^{k,+}$:

$$\begin{aligned} \mathbf{0}_{\mathbb{R}^k} &\in \partial_z(-\varphi)(\hat{x}, z) + \left\{ f(\hat{x}, \hat{y}) - \sum_{i=1}^k \eta_i \cdot \mathbf{e}_i + \mu \cdot \mathbf{e} \right\} \\ \eta_1, \dots, \eta_k &\geq 0 \quad \mu \in \mathbb{R} \quad \eta_1 \cdot (z^\top \mathbf{e}_1) = \dots = \eta_k \cdot (z^\top \mathbf{e}_k) = 0. \end{aligned} \quad (16)$$

Since $(-\varphi)(x, \cdot)$ is convex, we have $\partial_z(-\varphi)(x, z) = -\partial_z^c \varphi(x, z)$ provided φ is locally Lipschitz continuous at (x, z) . Therein, ∂ denotes the subdifferential in the sense of convex analysis while ∂^c represents the subdifferential in the sense of Clarke (cf. [7]). In literature (cf. [3, 8, 11, 20, 21] or [22]) there exist a lot of results that provide conditions ensuring different continuity properties (lower semicontinuity, continuity, Lipschitz continuity) of optimal-value-functions. Hence, postulating some conditions on (4) it should be possible to apply a concept of generalized differentiation (cf. [7] or [20]) to ψ . Without mentioning an exact proof the following result is likely to hold (one may apply the achievements from [22] to (15)): Choose $(\hat{x}, \hat{y}) \in X \times Y$ such that $\psi(\hat{x}, \hat{y}) = 0$ is valid. Assume that φ is locally Lipschitz continuous at any point (\hat{x}, z) where $z \in \mathbb{S}^{k,+}$ is satisfied. Then ψ is locally Lipschitz continuous at (\hat{x}, \hat{y}) (cf. [11]) and the following approximation formula holds true:

$$\begin{aligned} \partial^c \psi(\hat{x}, \hat{y}) &\subseteq \text{cl}^* \left(\text{conv} \left(\left\{ \left(\langle z, f \rangle'_x(\hat{x}, \hat{y}, z) - \xi, \langle z, f \rangle'_y(\hat{x}, \hat{y}, z) \right) \in X^* \times Y^* \mid \right. \right. \right. \\ &\quad \left. \left. \left. \xi \in \partial_x^c \varphi(\hat{x}, z), z \in \mathbb{S}^{k,+} \text{ satisfies (16)} \right\} \right) \right). \end{aligned} \quad (17)$$

Therein, the mapping $\langle z, f \rangle: X \times Y \times \mathbb{R}^k \longrightarrow \mathbb{R}$ is defined as

$$\forall x \in X \forall y \in Y \forall z \in \mathbb{R}^k: \quad \langle z, f \rangle(x, y, z) := z^\top f(x, y)$$

and, obviously, continuously Fréchet differentiable w.r.t. x and y provided f possesses this property. One may use another approximation formula now to find upper estimates for the partial Clarke subdifferentials of φ to make (17) applicable. This procedure seems promising in view of the formulation of necessary optimality conditions for the original semivectorial bilevel programming problem provided the weak* closure in (17) can be dropped. Note that due to the nature of bilevel programming the problem (14) is likely to be highly irregular (cf. [8, 13], and [28] where the authors show that common constraint qualifications of non-smooth optimization fail to hold at any feasible point of the optimal-value-reformulation of a finite-dimensional bilevel programming problem). Hence, in order to derive any necessary optimality conditions of KKT-type, one may have to make use of weak constraint qualifications like the *Weak-Basic-Constraint-Qualification*, presented in [12] for finite-dimensional bilevel programming problems, or *partial calmness* (cf. [27] and [28]).

Another possible way to reformulate (7) as a single-level problem seems to be the replacement of the lower level problem by some kind of optimality condition. Note that due to the convexity of the lower level problem (4) w.r.t. y some KKT-type optimality conditions are sufficient for optimality if only $z \in \mathbb{S}^{k,+}$ is satisfied. Postulating a regularity condition and some differentiability property to hold at any point $(x, y) \in \text{graph}(\Gamma)$, it is well-known (cf. [16]) that the KKT-conditions pose a necessary optimality condition for any point $(x, z, y) \in \text{graph}(\Psi)$. If we assume the lower level mappings f and g to be continuously Fréchet differentiable w.r.t. y , then these KKT-conditions take the following form:

$$\mathfrak{o}_{Y^*} = \langle z, f \rangle'_y(x, y, z) + g'_y(x, y) \circ \lambda \quad \langle g(x, y), \lambda \rangle = 0 \quad \lambda \in K_l^D \quad g(x, y) \in -K_l. \quad (18)$$

Consequently, using (18) one may replace (7) by

$$\begin{aligned} F(x, y) &\rightarrow \min_{x, y, z, \lambda} \\ G(x) &\in -K_u \\ -z &\in -\mathbb{R}_0^{k,+} \\ z^\top \mathbf{e} &= 1 \\ \langle z, f \rangle'_y(x, y, z) + g'_y(x, y) \circ \lambda &= \mathfrak{o}_{Y^*} \\ \langle g(x, y), \lambda \rangle &= 0 \\ -\lambda &\in -K_l^D \\ g(x, y) &\in -K_l \end{aligned} \quad (19)$$

provided any point $(x, y) \in \text{graph}(\Gamma)$ is partially regular w.r.t. y (cf. Lemma 2.2). Problem (19) is known as the *KKT-reformulation* of problem (7). It is shown in [9] that the introduction of Lagrange multipliers as new variables within the KKT-reformulation of a scalar finite-dimensional bilevel programming problem may lead to an optimization problem possessing local optimal solutions which are no local optimal solutions of the original bilevel programming problem. Furthermore, if the lower level feasible set of such a scalar finite-dimensional bilevel programming problem is given by inequality constraints, then the arising KKT-reformulation turns out to be a *mathematical program with complementarity constraints*. It is well-known from [24] that common constraint qualifications like MFCQ fail at any feasible point of such a problem. In our setting problem (19) is likely to face the same difficulties as described above. Hence, we have to scrutinize this reformulation approach critically.

A comparatively new idea how to deal with bilevel programming problems is described in [29] where the authors investigate a combination of optimal-value- and KKT-reformulation for scalar finite-dimensional bilevel programming problems. Applied to (7) this idea would lead to the surrogate problem (19) equipped with the additional optimal-value-constraint $z^\top f(x, y) - \varphi(x, z) \in -\mathbb{R}_0^+$. If the lower level mappings f and g are continuously Fréchet differentiable w.r.t. y and any lower level feasible point is partially regular w.r.t. y , then the arising problem is locally and globally equivalent to (7). However, this new surrogate problem faces the same problems concerning differentiability and/or regularity as presented by means of (12) and (19).

In the following section we want to exploit some kind of KKT-reformulation of simple semivectorial bilevel optimal control problems where some of the difficulties, described earlier, do not appear.

3 A simple semivectorial bilevel optimal control problem

Now we are going to study simple semivectorial bilevel optimal control problems whose structure allows the application of the scalarization approach presented in the previous section. We will attack the scalarized bilevel problem using the KKT-approach from bilevel programming to derive a surrogate single-level problem which is not only globally but also locally equivalent to the corresponding scalarized bilevel programming problem under some additional assumption. Moreover, we study the relationship of the original semivectorial bilevel optimal control problem and the aforementioned single-level problem. Afterwards, we construct a rather weak necessary optimality condition of Pontryagin-type for the original semivectorial bilevel optimal control problem applying the results of Vinter (cf. [25]). Let us start by introducing the problem. We are going to consider

$$\begin{aligned}
F_0(x(T), y(T)) &\rightarrow \min_{x, u, y, v} \\
G_0(t, x(t), u(t)) &= \dot{x}(t) \quad \text{a.e. } t \in [0, T] \\
x(0) &= x_0 \\
K(x(T)) &\leq \mathbf{0}_{\mathbb{R}^r} \\
H(x(T)) &= \mathbf{0}_{\mathbb{R}^s} \\
(y, v) &\in \Psi_{we}(x, u)
\end{aligned} \tag{20}$$

whose corresponding lower level problem is given as follows:

$$\begin{aligned}
\left. \begin{array}{l} f_1(x(T), y(T)) \\ \vdots \\ f_k(x(T), y(T)) \end{array} \right\} &\rightarrow \min_{y, v} \\
Ay(t) + Bv(t) + Mu(t) &= \dot{y}(t) \quad \text{a.e. } t \in [0, T] \\
y(0) &= y_0 \\
h(x(T), y(T)) &= \mathbf{0}_{\mathbb{R}^l}.
\end{aligned} \tag{21}$$

Forthwith, we assume that (20) possesses the properties stated below:

- $F_0: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$, $G_0: \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}^n$, $K: \mathbb{R}^n \rightarrow \mathbb{R}^r$, and $H: \mathbb{R}^n \rightarrow \mathbb{R}^s$ are continuously differentiable while $f_1, \dots, f_k: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ and $h: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^l$ are twice continuously differentiable,
- for any $x \in \mathbb{R}^n$ the mappings $f_1(x, \cdot), \dots, f_k(x, \cdot)$ are convex while $h(x, \cdot)$ is an affine operator,
- the function $f: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^k$ is defined as the vector function possessing the components f_1, \dots, f_k ,
- h possesses the component mappings $h_1, \dots, h_l: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$,
- the matrices $A \in \mathbb{R}^{m \times m}$, $B \in \mathbb{R}^{m \times q}$, $M \in \mathbb{R}^{m \times p}$, $x_0 \in \mathbb{R}^n$, and $y_0 \in \mathbb{R}^m$ are fixed,
- the follower's controllability matrix $\mathcal{S} := [B \mid AB \mid \dots \mid A^{m-1}B] \in \mathbb{R}^{m \times mq}$ possesses full row rank m ,
- the final time $T > 0$ is fixed,
- the minimization in (21) happens w.r.t. the ordering cone $\mathbb{R}_0^{k,+}$,
- the state functions x and y come from the Sobolev spaces $W_{1,1}^n[0, T]$ and $W_{1,1}^m[0, T]$, respectively,
- the control functions u and v are measurable functions from $L_1^p[0, T]$ and $L_1^q[0, T]$, respectively.

Let us define the Banach spaces E_1 , E_2 , E_s and E as follows:

$$E_1 := W_{1,1}^n[0, T] \times L_1^p[0, T] \quad E_2 := W_{1,1}^m[0, T] \times L_1^q[0, T] \quad E_s := W_{1,1}^n[0, T] \times W_{1,1}^m[0, T] \quad E := E_1 \times E_2.$$

Since the control functions u and v in the above optimal control problem (20) do not influence the objective values of leader and follower directly, it seems justifiable to consider local $W_{1,1}$ -minimizers. In the general terminology of optimal control local $W_{1,1}$ -minimizers are defined as local optimal solutions of an optimal control problem where the distance of the reference point to another feasible point is only computed w.r.t. the state variables, i.e. the distance of the corresponding control variables is disregarded. Transferring this concept to problem (20) we can formulate the following definition.

Definition 3.1

A point $(\hat{x}, \hat{u}, \hat{y}, \hat{v}) \in E$ is called **local $W_{1,1}$ -minimizer** of (20) if it is feasible for that problem and there exists $\varepsilon > 0$ such that for any feasible point $(x, u, y, v) \in E$ of (20) which satisfies $(x, y) \in \mathbb{U}_{E_s}^\varepsilon(\hat{x}, \hat{y})$ the inequality $F_0(\hat{x}(T), \hat{y}(T)) \leq F_0(x(T), y(T))$ holds.

It easily follows from the above definition that any local $W_{1,1}$ -minimizer of (20) is a local optimal solution of the latter problem. The opposite statement is not true in general. Hence, we are looking for a special class of local minimizers. Note that any global optimal solution of (20) is a local $W_{1,1}$ -minimizer of that problem. While considering the multiobjective lower level problem (21) the concept of $W_{1,1}$ -minimizers is redundant: when searching for weakly efficient points no distances have to be considered (cf. Definition 2.1). Furthermore, replacing (21) with its scalarized counterpart (4) and considering (7) demands to look just for global minimizers of the arising lower level problem (4). In fact, for any fixed $x \in X$ and $z \in \mathbb{S}^{k,+}$ the corresponding problem (4) is convex so that local minimizers, local $W_{1,1}$ -minimizers and global minimizers of this problem coincide.

The concept of defining local optimal solutions only w.r.t. a subset of all decision variables is not only popular in optimal control but also recently used in the paper [29] concerning finite-dimensional optimization. Therein, the authors consider a KKT-type reformulation of finite-dimensional optimistic bilevel programming problems but when analyzing local optimal solutions of the surrogate problem, they completely ignore the distance of the lower level Lagrange multipliers. Clearly, this way they can avoid the problems described in [9] since any (in the above sense) weakly local optimal solution of the surrogate problem corresponds to a (in common sense) local optimal solution of the bilevel programming problem under some appropriate convexity and regularity assumptions. However, this approach complicates the derivation of optimality conditions.

Following the ideas presented in Section 2, which are applicable due to the assumptions made on (21), we introduce a scalarization parameter $z \in \mathbb{S}^{k,+}$ and consider

$$\begin{aligned}
F_0(x(T), y(T)) &\rightarrow \min_{x, u, y, v, z} \\
G_0(t, x(t), u(t)) &= \dot{x}(t) && \text{a.e. } t \in [0, T] \\
x(0) &= x_0 \\
K(x(T)) &\leq \mathbf{0}_{\mathbb{R}^r} \\
H(x(T)) &= \mathbf{0}_{\mathbb{R}^s} \\
-z &\leq \mathbf{0}_{\mathbb{R}^k} \\
z^\top \mathbf{e} &= 1 \\
(y, v) &\in \Psi(x, u, z)
\end{aligned} \tag{22}$$

where $\Psi: E_1 \times \mathbb{S}^{k,+} \rightarrow \mathfrak{P}(E_2)$ is the solution-set-mapping of:

$$\begin{aligned}
z^\top f(x(T), y(T)) &\rightarrow \min_{y, v} \\
Ay(t) + Bv(t) + Mu(t) &= \dot{y}(t) && \text{a.e. } t \in [0, T] \\
y(0) &= y_0 \\
h(x(T), y(T)) &= \mathbf{0}_{\mathbb{R}^l}.
\end{aligned} \tag{23}$$

Forthwith, we stipulate to interpret the scalarization variable z as some kind of constant state function in order to clarify, how $W_{1,1}$ -minimizers of (22) are defined (i.e. we are considering the distance w.r.t. x , y and z). This way we might be able to deduce a slightly stronger optimality condition for problem (20) than we could derive when considering z as a control function. Similar as in Theorem 2.1 it can be shown that any local $W_{1,1}$ -minimizer $(\hat{x}, \hat{u}, \hat{y}, \hat{v}) \in E$ of (20) corresponds to a local $W_{1,1}$ -minimizer of (22). The opposite statement holds true under some additional assumptions comprising the closedness of Ψ at the reference points (\hat{x}, \hat{u}, z) where $z \in \mathbb{S}^{k,+}$ holds. The following remark presents a condition ensuring at least this property.

Remark 3.1

Let $(\hat{x}, \hat{u}, \hat{y}, \hat{v}) \in E$ be a feasible point of (20) such that the operator $h(\hat{x}(T), \cdot)$ is surjective. Then Ψ is closed at (\hat{x}, \hat{u}, z) where $z \in \mathbb{S}^{k,+}$ is arbitrarily chosen.

Proof We want to apply Lemma 2.2 for the upcoming proof. Therefor, let us specify $X := E_1$, $Y := E_2$, $Z_l := W_{1,1}^m[0, T] \times \mathbb{R}^l$, $K_l := \{\mathbf{o}_{Z_l}\}$ and define $g: X \times Y \rightarrow Z_l$ as follows for $((x, u), (y, v)) \in E_1 \times E_2$:

$$g(x, u, y, v) := \left(y(\cdot) - y_0 - \int_0^\cdot [Ay(s) + Bv(s) + Mu(s)] ds, h(x(T), y(T)) \right).$$

Obviously, for fixed (x, u) the mapping $g(x, u, \cdot, \cdot)$ is a bounded affine operator and hence, continuously Fréchet differentiable w.r.t. (y, v) . Its partial Fréchet derivative w.r.t. (y, v) at $(\hat{x}, \hat{u}, \tilde{y}, \tilde{v})$ for arbitrary points and directions $(\tilde{y}, \tilde{v}), (d_y, d_v) \in E_2$ is given below:

$$\begin{aligned} g'_{(y,v)}(\hat{x}, \hat{u}, \tilde{y}, \tilde{v})[(d_y, d_v)] &= g(\hat{x}, \hat{u}, d_y, d_v) - g(\hat{x}, \hat{u}, \mathbf{o}_{W_{1,1}^m[0,T]}, \mathbf{o}_{L_1^q[0,T]}) \\ &= \left(d_y(\cdot) - \int_0^\cdot [Ad_y(s) + Bd_v(s)] ds, \underbrace{h(\hat{x}(T), d_y(T)) - h(\hat{x}(T), \mathbf{o}_{\mathbb{R}^m})}_{=\nabla_y h(\hat{x}(T), y(T))d_y(T)} \right). \end{aligned}$$

We proceed showing that $g'_{(y,v)}(\hat{x}, \hat{u}, \tilde{y}, \tilde{v})$ is surjective (for any choice of $(\tilde{y}, \tilde{v}) \in E_2$). Therefor, choose an arbitrary point $(\omega, \xi) \in Z_l$. Due to the surjectivity of $h(\hat{x}(T), \cdot)$ we find $\zeta \in \mathbb{R}^m$ which is a solution of the linear system $h(\hat{x}(T), \zeta) = h(\hat{x}(T), \mathbf{o}_{\mathbb{R}^m}) + \xi$. Now consider the following *Volterra equation of second kind*:

$$\omega(\cdot) = d_y(\cdot) - \int_0^\cdot Ad_y(s) ds.$$

It possesses a solution $d^\omega \in W_{1,1}^m[0, T]$ (cf. [16]). Moreover, due to the postulated full rank of the controllability matrix \mathcal{S} we find a control function $\hat{d}_v \in L_1^q[0, T]$ such that there exists a solution $\tilde{d} \in W_{1,1}^m[0, T]$ of the following boundary value problem (cf. [4]):

$$\dot{d}_y(t) = Ad_y(t) + B\hat{d}_v(t), \quad d_y(0) = \mathbf{o}_{\mathbb{R}^m}, \quad d_y(T) = \zeta - d^\omega(T).$$

Now define $\hat{d}_y := \tilde{d} + d^\omega$. One can easily check that (\hat{d}_y, \hat{d}_v) solves $g'_{(y,v)}(\hat{x}, \hat{u}, \tilde{y}, \tilde{v})[(d_y, d_v)] = (\omega, \xi)$. Hence, $g'_{(y,v)}(\hat{x}, \hat{u}, \tilde{y}, \tilde{v})$ is surjective for any choice of $(\tilde{y}, \tilde{v}) \in E_2$ and consequently, all the points $(\hat{x}, \hat{u}, \tilde{y}, \tilde{v}) \in E$ are partially regular w.r.t. (y, v) . Now the statement of the remark follows from Lemma 2.2. \square

We want to take a closer look on problem (23) and characterise its optimal solutions by means of a multiplier-type optimality condition, i.e. we want to express $\Psi(x, u, z)$ explicitly. Therefor, we make use of a result by Vinter (cf. [25]).

Lemma 3.1

Choose an arbitrary point $(\hat{x}, \hat{u}, \hat{z}) \in E_1 \times \mathbb{S}^{k,+}$. For any feasible point $(\hat{y}, \hat{v}) \in E_2$ of (23) for fixed parameter $(\hat{x}, \hat{u}, \hat{z})$ the condition $(\hat{y}, \hat{v}) \in \Psi(\hat{x}, \hat{u}, \hat{z})$ is satisfied if and only if there exists a solution $\lambda \in \mathbb{R}^l$ of the following linear system:

$$\mathbf{o}_{\mathbb{R}^m} = \nabla_y f(\hat{x}(T), \hat{y}(T))^\top \hat{z} + \nabla_y h(\hat{x}(T), \hat{y}(T))^\top \lambda. \quad (24)$$

Moreover, if $h(\hat{x}(T), \cdot)$ is surjective, in the case of existence λ is uniquely determined.

Proof

\implies : Let $(\hat{y}, \hat{v}) \in \Psi(\hat{x}, \hat{u}, \hat{z})$ be satisfied. Then (\hat{y}, \hat{v}) is a local $W_{1,1}$ -minimizer of (23). Due to Theorem

6.2.1 in [25] there exists a function $\phi \in W_{1,1}^m[0, T]$ and a scalar $\lambda_0 \in \mathbb{R}_0^+$ with $(\phi, \lambda_0) \neq (\mathbf{o}_{W_{1,1}^m[0, T]}, 0)$ such that the conditions

$$\begin{aligned} -\dot{\phi}(t) &= A^\top \phi(t) & \text{a.e. } t \in [0, T] \\ \mathbf{o}_{\mathbb{R}^q} &= B^\top \phi(t) & \text{a.e. } t \in [0, T] \\ -\phi(T) - \lambda_0 \cdot \nabla_y f(\hat{x}(T), \hat{y}(T))^\top \hat{z} &\in \mathcal{N}_V(\hat{y}(T)) \end{aligned}$$

are satisfied where $V := \{\zeta \in \mathbb{R}^m \mid h(\hat{x}(T), \zeta) = \mathbf{o}_{\mathbb{R}^l}\}$ holds true. Since $h(\hat{x}(T), \cdot)$ is an affine operator between finite-dimensional spaces, we have:

$$\mathcal{N}_V(\hat{y}(T)) = \left\{ \nabla_y h(\hat{x}(T), \hat{y}(T))^\top \xi \mid \xi \in \mathbb{R}^l \right\}.$$

That is why we find $\xi \in \mathbb{R}^l$ such that $-\phi(T) = \lambda_0 \cdot \nabla_y f(\hat{x}(T), \hat{y}(T))^\top \hat{z} + \nabla_y h(\hat{x}(T), \hat{y}(T))^\top \xi$ holds true. From the second of the above conditions we deduce $B^\top \dot{\phi}(t) = \mathbf{o}_{\mathbb{R}^q}$ for almost every $t \in [0, T]$. Using the differential equation on ϕ provides $(AB)^\top \phi(t) = B^\top A^\top \phi(t) = -B^\top \dot{\phi}(t) = \mathbf{o}_{\mathbb{R}^q}$ almost everywhere. We proceed similarly in order to show that

$$\forall j = 1, \dots, m: \quad (A^{j-1}B)^\top \phi(t) = \mathbf{o}_{\mathbb{R}^q} \quad \text{a.e. } t \in [0, T]$$

holds true. Hence, we have $\mathcal{S}^\top \phi(t) = \mathbf{o}_{\mathbb{R}^m}$ for almost every $t \in [0, T]$. Since the columns of \mathcal{S}^\top are linearly independent, ϕ vanishes almost everywhere. Due to the non-triviality condition on (ϕ, λ_0) we have $\lambda_0 > 0$ and:

$$\mathbf{o}_{\mathbb{R}^m} = \lambda_0 \cdot \nabla_y f(\hat{x}(T), \hat{y}(T))^\top \hat{z} + \nabla_y h(\hat{x}(T), \hat{y}(T))^\top \xi. \quad (25)$$

That means dividing (25) by λ_0 and defining $\lambda := \frac{1}{\lambda_0} \cdot \xi$ shows (24).

[\Leftarrow]: Suppose there is $\lambda \in \mathbb{R}^l$ such that (24) is satisfied while $(\hat{y}, \hat{v}) \in E_2$ is a feasible point to (23) for fixed parameters $(\hat{x}, \hat{u}, \hat{z})$. Let $(y, v) \in E_2$ be an arbitrary feasible point to (23) for fixed parameters $(\hat{x}, \hat{u}, \hat{z})$. From $\hat{z} \in \mathbb{R}_0^{k,+}$ the objective of the latter problem is convex and differentiable while $h(\hat{x}(T), \cdot)$ is an affine operator. That is why we have:

$$\begin{aligned} \hat{z}^\top f(\hat{x}(T), y(T)) &\geq \hat{z}^\top f(\hat{x}(T), \hat{y}(T)) + \hat{z}^\top \nabla_y f(\hat{x}(T), \hat{y}(T))(y(T) - \hat{y}(T)) \\ &= \hat{z}^\top f(\hat{x}(T), \hat{y}(T)) - \lambda^\top \nabla_y h(\hat{x}(T), \hat{y}(T))(y(T) - \hat{y}(T)) \\ &= \hat{z}^\top f(\hat{x}(T), \hat{y}(T)) - \lambda^\top (h(\hat{x}(T), y(T)) - h(\hat{x}(T), \hat{y}(T))) = \hat{z}^\top f(\hat{x}(T), \hat{y}(T)). \end{aligned}$$

Hence, we have $(\hat{y}, \hat{v}) \in \Psi(\hat{x}, \hat{u}, \hat{z})$.

Assume that $\lambda, \tilde{\lambda} \in \mathbb{R}^l$ solve (24) while $h(\hat{x}(T), \cdot)$ is surjective. Then $\nabla_y h(\hat{x}(T), \hat{y}(T))$ possesses full row rank l while we have $\nabla_y h(\hat{x}(T), \hat{y}(T))^\top (\lambda - \tilde{\lambda}) = \mathbf{o}_{\mathbb{R}^m}$. Since the columns of $\nabla_y h(\hat{x}(T), \hat{y}(T))^\top$ are linearly independent, we deduce $\lambda - \tilde{\lambda} = \mathbf{o}_{\mathbb{R}^l}$ i.e. $\lambda = \tilde{\lambda}$. Consequently, in the case of its existence λ is uniquely determined. \square

Lemma 3.1 allows us to replace the scalarized lower level problem (23) by its feasibility and optimality conditions. Hence, we may consider the optimization problem

$$\begin{aligned} F_0(x(T), y(T)) &\rightarrow \min_{x, y, u, v, z, \lambda} \\ G_0(t, x(t), u(t)) &= \dot{x}(t) & \text{a.e. } t \in [0, T] \\ Ay(t) + Bv(t) + Mu(t) &= \dot{y}(t) & \text{a.e. } t \in [0, T] \\ x(0) &= x_0 \\ y(0) &= y_0 \\ \nabla_y f(x(T), y(T))^\top z + \nabla_y h(x(T), y(T))^\top \lambda &= \mathbf{o}_{\mathbb{R}^m} \\ K(x(T)) &\leq \mathbf{o}_{\mathbb{R}^r} \\ H(x(T)) &= \mathbf{o}_{\mathbb{R}^s} \\ h(x(T), y(T)) &= \mathbf{o}_{\mathbb{R}^l} \\ -z &\leq \mathbf{o}_{\mathbb{R}^k} \\ z^\top \mathbf{e} &= 1 \end{aligned} \quad (26)$$

where z and λ are interpreted as some kind of constant state variables in view of the definition of local $W_{1,1}$ -minimizers for this problem (i.e. we consider distances of feasible points only w.r.t. x, y, z and λ when speaking about neighbourhoods of a reference point). We already know that (20) and (22) may not be equivalent in the sense of local $W_{1,1}$ -minimizers. As mentioned in Section 2 the transformation of (22) into (26) is likely to create $W_{1,1}$ -minimizers of (26) which are no local $W_{1,1}$ -minimizers of (22). The following lemma characterises the relationship between the problems (22) and (26) in more detail and provides a criterion which ensures at least the equivalence of these two problems w.r.t. local $W_{1,1}$ -minimizers.

Lemma 3.2

1. Let $(\hat{x}, \hat{u}, \hat{y}, \hat{v}, \hat{z}) \in E \times \mathbb{R}^k$ be a global optimal solution (a local $W_{1,1}$ -minimizer) of (22). Then for any $\lambda \in \mathbb{R}^l$ satisfying (24) the point $(\hat{x}, \hat{u}, \hat{y}, \hat{v}, \hat{z}, \lambda) \in E \times \mathbb{R}^k \times \mathbb{R}^l$ is a global optimal solution (a local $W_{1,1}$ -minimizer) of (26). Furthermore, at least one such λ exists.
2. Let $(\hat{x}, \hat{u}, \hat{y}, \hat{v}, \hat{z}, \hat{\lambda}) \in E \times \mathbb{R}^k \times \mathbb{R}^l$ be a global optimal solution of (26). Then $(\hat{x}, \hat{u}, \hat{y}, \hat{v}, \hat{z}) \in E \times \mathbb{R}^k$ is a global optimal solution of (22).
3. Let $(\hat{x}, \hat{u}, \hat{y}, \hat{v}, \hat{z}, \hat{\lambda}) \in E \times \mathbb{R}^k \times \mathbb{R}^l$ be a local $W_{1,1}$ -minimizer of (26) where $h(\hat{x}(T), \cdot)$ is surjective. Then $(\hat{x}, \hat{u}, \hat{y}, \hat{v}, \hat{z}) \in E \times \mathbb{R}^k$ is a local $W_{1,1}$ -minimizer of (22).

Proof In the following we equip all the mentioned Banach spaces with the norms introduced in Section 1 if not explicitly stated otherwise.

1. We only prove this statement for local $W_{1,1}$ -minimizers. One may choose $\varepsilon = +\infty$ to see that the same argumentation applies to global optimal solutions.

Suppose that there exists $\hat{\lambda} \in \mathbb{R}^l$ satisfying (24) such that $(\hat{x}, \hat{u}, \hat{y}, \hat{v}, \hat{z}, \hat{\lambda})$ is no local $W_{1,1}$ -minimizer of (26). Then we find sequences $(x_\nu) \subseteq W_{1,1}^n[0, T]$ converging to \hat{x} , $(u_\nu) \subseteq L_1^p[0, T]$, $(y_\nu) \subseteq W_{1,1}^m[0, T]$ converging to \hat{y} , $(v_\nu) \subseteq L_1^q[0, T]$, $(z_\nu) \subseteq \mathbb{R}^k$ converging to \hat{z} and (λ_ν) converging to $\hat{\lambda}$ such that $(x_\nu, u_\nu, y_\nu, v_\nu, z_\nu, \lambda_\nu)$ is feasible for (26) and satisfies $F_0(x_\nu(T), y_\nu(T)) < F_0(\hat{x}(T), \hat{y}(T))$ for any $\nu \in \mathbb{N}$. Due to Lemma 3.1 the point $(x_\nu, u_\nu, y_\nu, v_\nu, z_\nu)$ is feasible to (22) for any $\nu \in \mathbb{N}$. Hence, $(\hat{x}, \hat{u}, \hat{y}, \hat{v}, \hat{z})$ is no local $W_{1,1}$ -minimizer of (22), which is a contradiction. Note that Lemma 3.1 also guarantees the existence of at least one $\lambda \in \mathbb{R}^l$ which satisfies (24).

2. Suppose that $(\hat{x}, \hat{u}, \hat{y}, \hat{v}, \hat{z})$ is no global optimal solution of the scalarized bilevel optimal control problem (22). Then we find a point $(\tilde{x}, \tilde{u}, \tilde{y}, \tilde{v}, \tilde{z}) \in E \times \mathbb{R}^k$, which is feasible for (22), such that $F_0(\tilde{x}(T), \tilde{y}(T)) < F_0(\hat{x}(T), \hat{y}(T))$ holds true. Due to Lemma 3.1 we find $\tilde{\lambda} \in \mathbb{R}^l$ such that $(\tilde{x}, \tilde{u}, \tilde{y}, \tilde{v}, \tilde{z}, \tilde{\lambda})$ is feasible for (26). This contradicts the global optimality of $(\hat{x}, \hat{u}, \hat{y}, \hat{v}, \hat{z}, \hat{\lambda})$.

3. Suppose that $(\hat{x}, \hat{u}, \hat{y}, \hat{v}, \hat{z})$ is no local $W_{1,1}$ -minimizer of (22). Then we find sequences $(x_\nu) \subseteq W_{1,1}^n[0, T]$ converging to \hat{x} , $(u_\nu) \subseteq L_1^p[0, T]$, $(y_\nu) \subseteq W_{1,1}^m[0, T]$ converging to \hat{y} , $(v_\nu) \subseteq L_1^q[0, T]$ as well as a sequence $(z_\nu) \subseteq \mathbb{R}^k$ converging to \hat{z} such that $(x_\nu, u_\nu, y_\nu, v_\nu, z_\nu)$ is feasible to (22) and satisfies $F_0(x_\nu(T), y_\nu(T)) < F_0(\hat{x}(T), \hat{y}(T))$ for any $\nu \in \mathbb{N}$. From Lemma 3.1 we derive the existence of a sequence $(\lambda_\nu) \subseteq \mathbb{R}^l$ such that the points $(x_\nu, u_\nu, y_\nu, v_\nu, z_\nu, \lambda_\nu)$ are feasible for problem (26).

We proceed by showing that the sequence (λ_ν) is bounded. Therefor, we assume the contrary, i.e.

$\|\lambda_\nu\| \rightarrow \infty$. W.l.o.g. we assume $\lambda_\nu \neq \mathbf{0}_{\mathbb{R}^l}$ for all $\nu \in \mathbb{N}$. Hence, the sequence $\left(\frac{\lambda_\nu}{\|\lambda_\nu\|}\right)$ is well-defined and

bounded. Consequently, it possesses a convergent subsequence which (w.l.o.g.) we denote by $\left(\frac{\lambda_\nu}{\|\lambda_\nu\|}\right)$

again. Let $\tilde{\lambda}$ be its limit point. Then obviously $\tilde{\lambda} \neq \mathbf{0}_{\mathbb{R}^l}$ holds true.

From the Sobolev Embedding Theorem presented in Lemma 1.1 we derive the existence of constants $c_1, c_2 > 0$ such that we have

$$\|x_\nu(T) - \hat{x}(T)\| \leq \|\tilde{x}_\nu - \hat{x}\|_{C^n[0, T]} \leq c_1 \cdot \|x_\nu - \hat{x}\|_{W_{1,1}^n[0, T]}$$

$$\|y_\nu(T) - \hat{y}(T)\| \leq \|y_\nu - \hat{y}\|_{C^m[0, T]} \leq c_2 \cdot \|y_\nu - \hat{y}\|_{W_{1,1}^m[0, T]}$$

for any $\nu \in \mathbb{N}$. That is why, due to the convergence of (x_ν) and (y_ν) , we easily see that $(x_\nu(T)) \subseteq \mathbb{R}^n$ converges to $\hat{x}(T)$ while $(y_\nu(T)) \subseteq \mathbb{R}^m$ converges to $\hat{y}(T)$.

Since f and g are continuously differentiable, the above observation automatically implies that $(\nabla_y f(x_\nu(T), y_\nu(T))) \subseteq \mathbb{R}^{k \times m}$ converges to $\nabla_y f(\hat{x}(T), \hat{y}(T))$ while $(\nabla_y h(x_\nu(T), y_\nu(T))) \subseteq \mathbb{R}^{l \times m}$ converges to $\nabla_y h(\hat{x}(T), \hat{y}(T))$. Moreover, since (z_ν) converges to \hat{z} , the sequence $(\nabla_y f(x_\nu(T), y_\nu(T)))^\top z_\nu$

is bounded. That is why we have:

$$\mathfrak{o}_{\mathbb{R}^m} = \lim_{\nu \rightarrow \infty} \left(\frac{\nabla_y f(x_\nu(T), y_\nu(T))^\top z_\nu}{\|\lambda_\nu\|} + \frac{\nabla_y h(x_\nu(T), y_\nu(T))^\top \lambda_\nu}{\|\lambda_\nu\|} \right) = \nabla_y h(\hat{x}(T), \hat{y}(T))^\top \tilde{\lambda}.$$

The fact that $\nabla_y h(\hat{x}(T), \hat{y}(T))^\top$ possesses linearly independent columns follows from the surjectivity of the affine operator $h(\hat{x}(T), \cdot)$ and implies $\tilde{\lambda} = \mathfrak{o}_{\mathbb{R}^l}$ which is a contradiction.

Hence, the sequence (λ_ν) is bounded and possesses a convergent subsequence, which (w.l.o.g.) we denote by (λ_ν) again. Let $\bar{\lambda} \in \mathbb{R}^l$ be its limit point. Reprising the above argumentation yields:

$$\begin{aligned} \mathfrak{o}_{\mathbb{R}^m} &= \lim_{\nu \rightarrow \infty} (\nabla_y f(x_\nu(T), y_\nu(T))^\top z_\nu + \nabla_y h(x_\nu(T), y_\nu(T))^\top \lambda_\nu) \\ &= \nabla_y f(\hat{x}(T), \hat{y}(T))^\top \hat{z} + \nabla_y h(\hat{x}(T), \hat{y}(T))^\top \bar{\lambda}. \end{aligned} \quad (27)$$

From the feasibility of $(\hat{x}, \hat{y}, \hat{u}, \hat{v}, \hat{z}, \hat{\lambda})$ for (26) we already know that

$$\mathfrak{o}_{\mathbb{R}^m} = \nabla_y f(\hat{x}(T), \hat{y}(T))^\top \hat{z} + \nabla_y h(\hat{x}(T), \hat{y}(T))^\top \hat{\lambda} \quad (28)$$

holds true. Combining (27) and (28) and recalling Lemma 3.1 we derive $\bar{\lambda} = \hat{\lambda}$ from the surjectivity of $h(\hat{x}(T), \cdot)$. This is a contradiction to the property of $(\hat{x}, \hat{y}, \hat{u}, \hat{v}, \hat{z}, \hat{\lambda})$ to be a local $W_{1,1}$ -minimizer of (26). \square

Now we are able to characterise the local and global relationship between the optimization problems (20) and (26) precisely. Therefor, we need the set-valued mapping $\Theta: E \rightarrow \mathfrak{P}(\mathbb{R}^k)$ defined as in Section 2:

$$\forall (x, u, y, v) \in E: \quad \Theta(x, u, y, v) := \{z \in \mathbb{S}^{k,+} \mid (y, v) \in \Psi(x, u, z)\}.$$

Theorem 3.1

1. Let $(\hat{x}, \hat{u}, \hat{y}, \hat{v}) \in E$ be a global optimal solution (a local $W_{1,1}$ -minimizer) of (20). Then for any $\hat{z} \in \Theta(\hat{x}, \hat{u}, \hat{y}, \hat{v})$ and any $\lambda \in \mathbb{R}^l$ satisfying (24) the point $(\hat{x}, \hat{u}, \hat{y}, \hat{v}, \hat{z}, \lambda) \in E \times \mathbb{R}^k \times \mathbb{R}^l$ is a global optimal solution (a local $W_{1,1}$ -minimizer) of (26). Furthermore, there exists at least one $\hat{z} \in \Theta(\hat{x}, \hat{u}, \hat{y}, \hat{v})$ and for any such \hat{z} we can find a vector λ satisfying (24).
2. Let $(\hat{x}, \hat{u}, \hat{y}, \hat{v}, \hat{z}, \hat{\lambda}) \in E \times \mathbb{R}^k \times \mathbb{R}^l$ be a global optimal solution of (26). Then $(\hat{x}, \hat{u}, \hat{y}, \hat{v}) \in E$ is a global optimal solution of (20).
3. Assume that for any $z \in \Theta(\hat{x}, \hat{u}, \hat{y}, \hat{v})$ there exists $\lambda \in \mathbb{R}^l$ such that $(\hat{x}, \hat{u}, \hat{y}, \hat{v}, z, \lambda) \in E \times \mathbb{R}^k \times \mathbb{R}^l$ is a local $W_{1,1}$ -minimizer of (26). Furthermore, let $h(\hat{x}(T), \cdot)$ be surjective. Then $(\hat{x}, \hat{u}, \hat{y}, \hat{v}) \in E$ is a local $W_{1,1}$ -minimizer of (20).

Proof The proof is straightforward applying Theorem 2.1, Remark 3.1, and Lemma 3.2. \square

As mentioned earlier we have to interpret z and λ as constant state variables of (26) in order to clarify the meaning of local $W_{1,1}$ -minimizers of that problem. Hence, we can reformulate (26) equivalently as a real optimal control problem. This is necessary if we want to apply the results in [25] in order to derive an optimality condition for (20). Therefore, consider the following optimal control problem (for simplicity

we keep the notations for z and λ):

$$\begin{aligned}
F_0(x(T), y(T)) &\rightarrow \min_{x, u, y, v} \\
G_0(t, x(t), u(t)) &= \dot{x}(t) && \text{a.e. } t \in [0, T] \\
Ay(t) + Bv(t) + Mu(t) &= \dot{y}(t) && \text{a.e. } t \in [0, T] \\
\mathbf{0}_{\mathbb{R}^k} &= \dot{z}(t) && \text{a.e. } t \in [0, T] \\
\mathbf{0}_{\mathbb{R}^l} &= \dot{\lambda}(t) && \text{a.e. } t \in [0, T] \\
x(0) &= x_0 \\
y(0) &= y_0 \\
\nabla_y f(x(T), y(T))^\top z(T) + \nabla_y h(x(T), y(T))^\top \lambda(T) &= \mathbf{0}_{\mathbb{R}^m} \\
K(x(T)) &\leq \mathbf{0}_{\mathbb{R}^r} \\
H(x(T)) &= \mathbf{0}_{\mathbb{R}^s} \\
h(x(T), y(T)) &= \mathbf{0}_{\mathbb{R}^l} \\
-z(T) &\leq \mathbf{0}_{\mathbb{R}^k} \\
z(T)^\top \mathbf{e} &= 1.
\end{aligned} \tag{29}$$

The above optimal control problem possesses the state variables x , y , z , and λ while u and v are the controls. Naturally, for the purpose of completeness, we claim $z \in W_{1,1}^k[0, T]$ and $\lambda \in W_{1,1}^l[0, T]$. We may apply Lemma 3.2 in order to see that this problem is equivalent to (22) in the sense of local $W_{1,1}$ -minimizers and global optimal solutions provided the affine operator $h(x(T), \cdot)$ is surjective for arbitrary $x \in W_{1,1}^n[0, T]$.

Finally, the following theorem presents a weak necessary optimality condition for problem (20) without assuming any additional constraint qualification.

Theorem 3.2 (*Maximum Principle*)

Let $(\hat{x}, \hat{u}, \hat{y}, \hat{v}) \in E$ be a local $W_{1,1}$ -minimizer of (20). Then we find a function $\phi \in W_{1,1}^n[0, T]$, vectors $\hat{z} \in \mathbb{S}^{k,+}$, $\hat{\lambda} \in \mathbb{R}^l$, $\alpha \in \mathbb{R}^m$, $\beta \in \mathbb{R}_0^{r,+}$, $\gamma \in \mathbb{R}^s$, $\delta \in \mathbb{R}^l$ as well as $\eta \in \mathbb{R}_0^{k,+}$ and scalars $\sigma \in \mathbb{R}$ as well as $\tau \geq 0$ such that the following conditions hold (for almost every $t \in [0, T]$):

1. *Non-triviality condition*

$$\|\phi\|_{W_{1,1}^n[0, T]} + \|\alpha\|_\infty + \|\beta\|_\infty + \|\gamma\|_\infty + \|\delta\|_\infty + \|\eta\|_\infty + |\sigma| + \tau > 0,$$

2. *Adjoint condition*

$$-\dot{\phi}(t)^\top = \phi(t)^\top \nabla_x G_0(t, \hat{x}(t), \hat{u}(t)),$$

3. *Weierstrass condition*

$$\phi(t)^\top G_0(t, \hat{x}(t), \hat{u}(t)) = \max_{\mathbf{u} \in \mathbb{R}^p} \phi(t)^\top G_0(t, \hat{x}(t), \mathbf{u}),$$

4. *Transversality condition*

$$\begin{aligned}
-\phi(T) &= \tau \cdot \nabla_x F_0(\hat{x}(T), \hat{y}(T))^\top + \sum_{i=1}^k (\hat{z}^\top \mathbf{e}_i) \nabla_{yy}^2 f_i(\hat{x}(T), \hat{y}(T))^\top \alpha + \sum_{j=1}^l (\hat{\lambda}^\top \mathbf{e}_j) \nabla_{yx}^2 h_j(\hat{x}(T), \hat{y}(T))^\top \alpha \\
&\quad + \nabla K(\hat{x}(T))^\top \beta + \nabla H(\hat{x}(T))^\top \gamma + \nabla_x h(\hat{x}(T), \hat{y}(T))^\top \delta,
\end{aligned}$$

5. *Multiplier conditions*

$$\begin{aligned}
\mathbf{0}_{\mathbb{R}^m} &= \tau \cdot \nabla_y F_0(\hat{x}(T), \hat{y}(T))^\top + \sum_{i=1}^k (\hat{z}^\top \mathbf{e}_i) \nabla_{yy}^2 f_i(\hat{x}(T), \hat{y}(T))^\top \alpha + \nabla_y h(\hat{x}(T), \hat{y}(T))^\top \delta \\
\mathbf{0}_{\mathbb{R}^k} &= \nabla_y f(\hat{x}(T), \hat{y}(T))^\top \alpha - \eta + \sigma \cdot \mathbf{e} \\
\mathbf{0}_{\mathbb{R}^l} &= \nabla_y h(\hat{x}(T), \hat{y}(T))^\top \alpha,
\end{aligned}$$

6. Complementarity conditions

$$K(\hat{x}(T))^\top \beta = 0 \quad \hat{z}^\top \eta = 0,$$

7. Lower level optimality condition

$$\mathbf{o}_{\mathbb{R}^m} = \nabla_y f(\hat{x}(T), \hat{y}(T))^\top \hat{z} + \nabla_y h(\hat{x}(T), \hat{y}(T))^\top \hat{\lambda}.$$

Proof As mentioned earlier we equip all the considered Banach spaces with the norms introduced in Section 1. Due to Theorem 3.1 we find $\hat{z} \in \mathbb{S}^{k,+}$ and $\hat{\lambda} \in \mathbb{R}^l$ such that $(\hat{x}, \hat{u}, \hat{y}, \hat{v}, \hat{z}, \hat{\lambda})$ is a local $W_{1,1}$ -minimizer of problem (26). Hence, the lower level optimality condition 7. holds. Interpreting \hat{z} and $\hat{\lambda}$ as constant functions easily shows that $(\hat{x}, \hat{u}, \hat{y}, \hat{v}, \hat{z}, \hat{\lambda})$ is a local $W_{1,1}$ -minimizer of (29). It is not difficult to check that the necessary optimality conditions of Theorem 6.2.3 in [25] are applicable to this problem. Hence, we find functions $\phi_x \in W_{1,1}^n[0, T]$, $\phi_y \in W_{1,1}^m[0, T]$, $\phi_z \in W_{1,1}^k[0, T]$, and $\phi_\lambda \in W_{1,1}^l[0, T]$, scalars $\sigma \in \mathbb{R}$ and $\tau \geq 0$ as well as vectors $\alpha \in \mathbb{R}^m$, $\beta \in \mathbb{R}_0^{r,+}$, $\gamma \in \mathbb{R}^s$, $\delta \in \mathbb{R}^l$, $\eta \in \mathbb{R}_0^{k,+}$, $\zeta \in \mathbb{R}^n$, and $\xi \in \mathbb{R}^m$ such that the following conditions hold (for almost every $t \in [0, T]$):

– Non-triviality condition

$$\|\phi_x\| + \|\phi_y\| + \|\phi_z\| + \|\phi_\lambda\| + \|\alpha\| + \|\beta\| + \|\gamma\| + \|\delta\| + \|\zeta\| + \|\eta\| + |\sigma| + \tau > 0, \quad (30)$$

– Adjoint conditions

$$\begin{aligned} -\dot{\phi}_x(t)^\top &= \phi_x(t)^\top \nabla_x G_0(t, \hat{x}(t), \hat{u}(t)) & -\dot{\phi}_y(t)^\top &= \phi_y(t)^\top A \\ -\dot{\phi}_z(t)^\top &= \mathbf{o}_{\mathbb{R}^k}^\top & -\dot{\phi}_\lambda(t)^\top &= \mathbf{o}_{\mathbb{R}^l}^\top, \end{aligned} \quad (31)$$

– Weierstrass conditions

$$\begin{aligned} \phi_x(t)^\top G_0(t, \hat{x}(t), \hat{u}(t)) + \phi_y(t)^\top M \hat{u}(t) &= \max_{\mathbf{u} \in \mathbb{R}^p} \{ \phi_x(t)^\top G_0(t, \hat{x}(t), \mathbf{u}) + \phi_y(t)^\top M \mathbf{u} \} \\ B^\top \phi_y(t) &= \mathbf{o}_{\mathbb{R}^q}, \end{aligned} \quad (32)$$

– Initial conditions

$$\phi_x(0) = \zeta \quad \phi_y(0) = \xi \quad \phi_z(0) = \mathbf{o}_{\mathbb{R}^k} \quad \phi_\lambda(0) = \mathbf{o}_{\mathbb{R}^l}, \quad (33)$$

– Terminal conditions

$$\begin{aligned} -\phi_x(T) &= \tau \cdot \nabla_x F_0(\hat{x}(T), \hat{y}(T))^\top + \sum_{i=1}^k (\hat{z}^\top \mathbf{e}_i) \nabla_{yx}^2 f_i(\hat{x}(T), \hat{y}(T))^\top \alpha \\ &\quad + \sum_{j=1}^l (\hat{\lambda}^\top \mathbf{e}_j) \nabla_{yx}^2 h_j(\hat{x}(T), \hat{y}(T))^\top \alpha + \nabla K(\hat{x}(T))^\top \beta \\ &\quad + \nabla H(\hat{x}(T))^\top \gamma + \nabla_x h(\hat{x}(T), \hat{y}(T))^\top \delta \end{aligned} \quad (34)$$

$$-\phi_y(T) = \tau \cdot \nabla_y F_0(\hat{x}(T), \hat{y}(T))^\top + \sum_{i=1}^k (\hat{z}^\top \mathbf{e}_i) \nabla_{yy}^2 f_i(\hat{x}(T), \hat{y}(T))^\top \alpha + \nabla_y h(\hat{x}(T), \hat{y}(T))^\top \delta$$

$$-\phi_z(T) = \nabla_y f(\hat{x}(T), \hat{y}(T)) \alpha - \eta + \sigma \cdot \mathbf{e}$$

$$-\phi_\lambda(T) = \nabla_y h(\hat{x}(T), \hat{y}(T)) \alpha,$$

– Complementarity conditions

$$K(\hat{x}(T))^\top \beta = 0 \quad \hat{z}^\top \eta = 0. \quad (35)$$

Note that the derivative $\nabla_{yy}^2 h(\hat{x}(T), \hat{y}(T))$ vanishes since $h(\hat{x}(T), \cdot)$ is an affine operator. From (31) and (33) we easily see that ϕ_z and ϕ_λ are equal to zero almost everywhere. Similar as in the proof of Lemma 3.1 we can show that ϕ_y vanishes as well. That causes $\xi = \mathbf{o}_{\mathbb{R}^m}$. Furthermore, note that we can drop the summand $\|\zeta\|$ from condition (30) since it is the initial value of ϕ_x and hence dominated by $\|\phi_x\|$ in the following sense: If $\|\phi_x\|$ is positive, then (30) holds. Otherwise, if $\|\phi_x\|$ equals zero, then $\zeta = \phi_x(0)$ vanishes as well. Hence, we finally rename $\phi := \phi_x$ to derive the optimality conditions 1. to 6. of the theorem. \square

Obviously, the above Maximum Principle is likely to be degenerated (i.e. it could happen that it holds for any feasible point of (20) or at least for a lot of feasible points of this problem) since we postulated no regularity condition on (29). Especially, the gradients of the boundary conditions may be (positively) linear dependent for numerous feasible points of that problem. This means, defining $\phi := \mathbf{o}_{W_{1,1}^n[0,T]}$ and $\tau := 0$, one could choose the multipliers (which correspond to the boundary conditions) not all equal to zero and satisfying the transversality condition, the multiplier condition, and the complementarity condition. In order to construct a constraint qualification for (29), one needs to formulate some restrictive assumptions on the dynamics $\dot{x}(t) = G_0(t, x(t), u(t))$ (controllability) and an MFCQ-type condition on the boundary constraints. Observe that MFCQ for problem (29) comprises the variables z and λ , which are not part of the original semivectorial bilevel programming problem. Hence, a constraint qualification on (29) is likely to depend on more than the initial data, which causes some additional difficulties. However, in this paper we are not going to deal with that question in more detail and leave it as a problem of our future research.

References

1. Adams, R., Fournier, J.: Sobolev Spaces. Elsevier Science, Oxford (2003)
2. Aubin, F., Frankowska, H.: Set-valued Analysis. Birkhäuser-Verlag, Basel, Berlin, Boston (1990)
3. Bank, B., Guddat, J., Klatte, D., Kummer, B., Tammer, K.: Non-Linear Parametric Optimization. Birkhäuser-Verlag, Basel, Boston, Stuttgart (1983)
4. Barnett, S., Cameron, R.G.: Introduction to Mathematical Control Theory. Oxford University Press, New York (1990)
5. Bonnans, J.F., Shapiro, A.: Perturbation Analysis of Optimization Problems. Springer-Verlag, New York, Berlin, Heidelberg (2000)
6. Bonnel, H., Morgan, J.: Optimality conditions for semivectorial bilevel convex optimal control problems. Computational and Analytical Mathematics **50**, 45–78 (2013)
7. Clarke, F.H.: Optimization and nonsmooth analysis. Wiley-Verlag, New York (1983)
8. Dempe, S.: Foundations of Bilevel Programming. Kluwer Academic Publishers, Dordrecht (2002)
9. Dempe, S., Dutta, J.: Is bilevel programming a special case of a mathematical program with complementarity constraints? Mathematical Programming, Series A **131**, 37–48 (2010)
10. Dempe, S., Gadhi, N., Zemkoho, A.B.: New optimality conditions for the semivectorial bilevel optimization problem. Journal of Optimization Theory and Applications **157**(1), 54–74 (2013)
11. Dempe, S., Mehlitz, P.: Lipschitz continuity of the optimal value function in parametric optimization. Journal of Global Optimization pp. 1–15 (2014). DOI 10.1007/s10898-014-0169-z
12. Dempe, S., Zemkoho, A.B.: The Generalized Mangasarian-Fromovitz Constraint Qualification and Optimality Conditions for Bilevel Programs. Journal of Optimization Theory and Applications **148**(1), 46–68 (2011)
13. Dempe, S., Zemkoho, A.B.: The bilevel programming problem: reformulations, constraint qualifications and optimality conditions. Mathematical Programming **138**, 447–473 (2013)
14. Ehrgott, M.: Multicriteria Optimization. Springer, Berlin, Heidelberg (2005)
15. Fiacco, A., Kyparisis, J.: Convexity and concavity properties of the optimal value function in nonlinear parametric programming. Journal of Optimization Theory and Applications **48**(1), 95–126 (1986)
16. Jahn, J.: Introduction to the Theory of Nonlinear Optimization. Springer-Verlag, Berlin, Heidelberg, New York (1996)
17. Jahn, J.: Vector Optimization. Springer-Verlag, Berlin, Heidelberg, New York (2004)
18. Kurcysz, S., Zowe, J.: Regularity and Stability for the Mathematical Programming Problem in Banach Spaces. Applied Mathematics and Optimization **5**, 49–62 (1979)
19. Liu, B., Wan, Z., Chen, J., Wang, G.: Optimality conditions for pessimistic semivectorial bilevel programming problems. Journal of Inequalities and Applications **41**(1) (2014). DOI 10.1186/1029-242X-2014-41
20. Mordukhovich, B.S.: Variational Analysis and Generalized Differentiation I and II. Springer-Verlag, Berlin, Heidelberg (2006)
21. Mordukhovich, B.S., Nam, N.M.: Variational stability and marginal functions via generalized differentiation. Mathematics of Operations Research **30**(4), 800–816 (2005)
22. Mordukhovich, B.S., Nam, N.M., Yen, N.D.: Subgradients of marginal functions in parametric mathematical programming. Mathematical Programming **116**, 369–396 (2009)
23. Robinson, S.: Stability theory for systems of inequalities, Part II: Differentiable nonlinear systems. SIAM Journal on Numerical Analysis **13**(4), 497–513 (1976)
24. Scheel, H., Scholtes, S.: Mathematical Programs with Complementarity Constraints: Stationarity, Optimality, and Sensitivity. Mathematics of Operations Research **25**(1), 1–22 (2000)
25. Vinter, R.: Optimal Control. Birkhäuser-Verlag, Boston, Basel, Berlin (2000)
26. Werner, D.: Funktionalanalysis. Springer-Verlag, Berlin, Heidelberg (1995)
27. Ye, J.J.: Optimal strategies for bilevel dynamic problems. SIAM Journal on Optimization and Control **35**(2), 512–531 (1997)
28. Ye, J.J., Zhu, D.L.: Optimality conditions for bilevel programming problems. Optimization **33**, 9–27 (1995)
29. Ye, J.J., Zhu, D.L.: New Necessary Optimality Conditions for Bilevel Programs by Combining the MPEC and Value Function Approaches. SIAM Journal on Optimization **20**(4), 1885–1905 (2010)