Weak and strong stationarity in generalized bilevel programming and bilevel optimal control
Patrick Mehlitz, Gerd Wachsmuth

Weak and strong stationarity in generalized bilevel programming and bilevel optimal control

TU Bergakademie Freiberg
Fakultät für Mathematik und Informatik
Prüferstraße 9
09596 FREIBERG
http://www.mathe.tu-freiberg.de
Weak and strong stationarity in generalized bilevel programming and bilevel optimal control

Patrick Mehlitz · Gerd Wachsmuth

Abstract In this article, we consider a general bilevel programming problem in reflexive Banach spaces with a convex lower level problem. In order to derive necessary optimality conditions for the bilevel problem, it is transferred to a mathematical program with complementarity constraints (MPCC). We introduce a notion of weak stationarity and exploit the concept of strong stationarity for MPCCs in reflexive Banach spaces, recently developed by the second author, and we apply these concepts to the reformulated bilevel programming problem. Constraint qualifications are presented, which ensure that local optimal solutions satisfy the weak and strong stationarity conditions. Finally, we discuss a certain bilevel optimal control problem by means of the developed theory. Its weak and strong stationarity conditions of Pontryagin-type and some controllability assumptions ensuring strong stationarity of any local optimal solution are presented.

Keywords Bilevel Programming · Programming in Banach Spaces · Mathematical Program with Complementarity Constraints · Stationarity · Bilevel Optimal Control

Mathematics Subject Classification (2000) 46N10 · 49K15 · 90C33 · 90C48

1 Introduction

It is well-known in optimization theory that mathematical programs with complementarity constraints (MPCCs) are difficult to handle since all constraint qualifications of reasonable strength fail to hold at any feasible point. This applies to both, the finite-dimensional and the infinite-dimensional situation. On the other hand, such problems arise frequently in many different applications. Hence, in the past a huge effort was put in constructing suitable regularity conditions, stationarity concepts, and relaxation techniques, which are applicable to MPCCs in theoretical and numerical procedures, cf. [16, 22, 24, 25, 28] and the references therein. Mainly, this analysis was done in finite-dimensional spaces. To the best of our knowledge, [26] is the first contribution which introduced a concept of strong stationarity applicable to a generalized MPCC in reflexive Banach spaces. The consideration of this abstract setting is rather important from the practical perspective since, e.g., there exist many dynamic optimization problems resulting in optimal control problems with complementarity constraints. Especially, when considering bilevel optimal control problems, it is possible to replace the original hierarchical optimization problem by a single-level optimal control problem which possesses a complementarity constraint provided the lower level problem is regular, convex, and equipped with pure state, mixed, or pure control inequality constraints.

After mentioning some notations and preliminary results in Section 2 we are going to recall the generalized concept of strong stationarity and introduce weak stationarity for MPCCs in Section 3. Afterwards,
we apply these concepts to a bilevel programming problem in Banach spaces whose lower level problem is regular and convex in Section 4. We present regularity conditions, which imply that these stationarity conditions hold at local minimizers of the original bilevel programming problem. Furthermore, we specify the results for finite-dimensional problems in order to show, that our concepts are an appropriate generalization. Section 5 is dedicated to the study of a certain bilevel optimal control problem with pure control inequality constraints in the lower level. After a brief introduction to bilevel optimal control we apply the aforementioned theory in order to derive necessary optimality conditions of Pontryagin-type for the latter problem. It will be shown that the controllability of a linear autonomous system is sufficient for any local minimizer of the given bilevel optimal control problem to be strongly stationary.

2 Notation and preliminary results

We use \( \mathbb{N}, \mathbb{N}^+, \mathbb{R}, \mathbb{R}^n, \mathbb{R}^{n^+}, \) and \( \mathbb{R}^{n \times m} \) to denote the natural numbers (with zero), the positive natural numbers, the real numbers, the set of all real vectors with \( n \) components, the cone of all vectors from \( \mathbb{R}^n \) possessing nonnegative components, and the set of all real matrices with \( n \) rows and \( m \) columns, respectively. For an arbitrary matrix \( \mathbf{Q} \in \mathbb{R}^{n \times m}, \) \( \mathbf{Q}^\top \) denotes its transpose. Furthermore, we use \( \mathbf{0} \in \mathbb{R}^n, \mathbf{O} \in \mathbb{R}^{n \times m}, \) and \( \mathbf{I} \in \mathbb{R}^{n \times n} \) to represent the zero vector, the zero matrix, and the identity matrix of appropriate dimensions, respectively.

Let \( X \) be a real reflexive Banach space with norm \( \| \cdot \| \) and zero vector \( \mathbf{0}_X \), and let \( A \subseteq X \) be a nonempty subset of \( X \). We denote by \( \text{lin}(A), \text{cone}(A), \text{conv}(A), \) and \( \text{cl}(A) \) the smallest subspace of \( X \) containing \( A \), the conic hull of \( A \), the convex hull of \( A \), and the closure of \( A \), respectively. The (topological) dual space of \( X \) is denoted by \( X^* \) and \( \langle \cdot, \cdot \rangle : X \times X^* \to \mathbb{R} \) is the corresponding dual pairing. Let \( B \subseteq X^* \) be nonempty as well. Then we define the polar cones and annihilators of the sets \( A \) and \( B \) by

\[
\begin{align*}
A^o & := \{ x^* \in X^* \mid \forall x \in A : \langle x, x^* \rangle \leq 0 \} \quad & A^+ & := \{ x^* \in X^* \mid \forall x \in A : \langle x, x^* \rangle = 0 \} \\
B^o & := \{ x \in X \mid \forall x^* \in B : \langle x, x^* \rangle \leq 0 \} \quad & B^+ & := \{ x \in X \mid \forall x^* \in B : \langle x, x^* \rangle = 0 \},
\end{align*}
\]

respectively. Note that due to the reflexivity of \( X \), the above notation is consistent. For the purpose of simplicity, we leave curly brackets away when considering singletons, i.e. we set \( \text{lin}(x) := \text{lin}\{x\} \), \( \text{cone}(x) := \text{cone}\{x\} \), \( x^o := \{x\}^o \), and \( x^+ := \{x\}^+ \) for any \( x \in X \) and similar definitions shall hold for all \( x^* \in X^* \). For a closed, convex set \( C \subseteq X \) and a fixed vector \( x \in C \), we define the cone of feasible directions and the Bouligand tangent cone to \( C \) at \( x \) by

\[
R_C(x) := \text{cone}(C - \{x\}) \quad \text{and} \quad T_C(x) := \text{cl}(R_C(x)),
\]

respectively. Moreover, if \( x^* \in T_C(x)^o \) is chosen, then the critical cone to \( C \) with respect to (w.r.t.) \( x \) and \( x^* \) is defined by

\[
K_C(x,x^*) := T_C(x) \cap (x^*)^o.
\]

In case that \( C \) is additionally a cone, we have

\[
R_C(x) = C + \text{lin}(x) \quad T_C(x) = \text{cl}(C + \text{lin}(x)) \quad T_C(x)^o = C^o \cap x^+.
\]

(1)

The next lemma summarizes some properties of the above operators.

**Lemma 2.1** Let \( K_1, K_2 \subseteq X \) be nonempty, closed, convex cones. Then we have

\[
\begin{align*}
(K_1^o)^o &= K_1 & (K_1 + K_2)^o &= K_1^o \cap K_2^o \\
(K_1 \cap K_2)^o &= \text{cl}(K_1^o + K_2^o) & (K_1^o)^+ &= K_1 \cap (-K_1) \\
K_1^+ &= K_1^o \cap (-K_1)^o.
\end{align*}
\]
Proof For the statements (2a), (2b), and (2c), we refer to [5, Proposition 2.40, (2.31), and (2.32)], respectively. Using (2a), we obtain (2d) by

\[(K^o)^{\perp} = \{ x \in X \mid \forall x^* \in K^o_1 : \langle x, x^* \rangle = 0 \} = \{ x \in X \mid \forall x^* \in K^o_1 : \langle x, x^* \rangle \leq 0 \} \cap \{ x \in X \mid \forall x^* \in -K^o_1 : \langle x, x^* \rangle \leq 0 \} = (K^o_1)^{\perp} \cap (-K^o_1)^{\perp} = K_1 \cap (-K_1).\]

A similar argument yields (2e). \hfill \square

Additionally, observe that \( \text{lin}(x) = (x^+)^{\perp} = (x^o)^c \) holds for an arbitrary point \( x \in X \).

We stipulate that throughout the whole paper the intersection \( \cap \) is a stronger operation than the Minkowski sum, i.e. \( S_1 + S_2 \cap S_3 \) is to be read as \( S_1 + (S_2 \cap S_3) \), for arbitrary subsets \( S_1, S_2, \) and \( S_3 \) of a linear space.

By a simple calculation, we obtain the following lemma. Its proof is straightforward and, hence, omitted.

Lemma 2.2 Let \( A \subseteq X \) be a nonempty set, \( U \subseteq X \) be a linear subspace of \( X \), and \( \xi \in U \). Then we have

\[ (A + \text{lin}(\xi)) \cap U = A \cap U + \text{lin}(\xi). \]

Let \( U, V, \) and \( W \) be arbitrary Banach spaces. By \( L[U, V] \) we denote the space of bounded linear operators mapping from \( U \) to \( V \). For any operator \( F \in L[U, V] \), the operator \( F^* \in L[V^*, U^*] \) denotes the adjoint of \( F \).

Suppose that \( \theta: U \times V \longrightarrow W \) is a twice Fréchet differentiable mapping and \( (\hat{u}, \hat{v}) \in U \times V \) is fixed. Then the linear operators \( \theta'(\hat{u}, \hat{v}) \in L[U \times V, W], \theta'_\omega(\hat{u}, \hat{v}) \in L[U, W], \theta'(2)(\hat{u}, \hat{v}) \in L[U \times V, L[U \times V, W]], \theta'(2)_\omega(\hat{u}, \hat{v}) \in L[L[U, L[U, W]] \times L[U, L[U, W]]], \) and \( \theta'(2)_\omega(\hat{u}, \hat{v}) \in L[V, L[U, W]] \) denote its first order Fréchet derivative, partial first order Fréchet derivative w.r.t. \( u \), second order Fréchet derivative, partial second order Fréchet derivative w.r.t. \( u \) and \( v \) at \( (\hat{u}, \hat{v}) \), respectively. As it is common in literature, we will identify any second order Fréchet derivative with a bounded bilinear form.

Suppose \( U = \mathbb{R}^n, V = \mathbb{R}^m \) as well as \( W = \mathbb{R}^k \), choose an arbitrary point \( (\hat{u}, \hat{v}) \in \mathbb{R}^n \times \mathbb{R}^m \), and let \( \vartheta: \mathbb{R}^n \times \mathbb{R}^m \longrightarrow \mathbb{R}^k \) be differentiable. Then the real matrices \( \nabla \vartheta(\hat{u}, \hat{v}) \in \mathbb{R}^{k \times (n+m)} \) and \( \nabla_u \vartheta(\hat{u}, \hat{v}) \in \mathbb{R}^{k \times n} \) denote the Jacobian and the partial Jacobian w.r.t. \( u \) of \( \vartheta \) at \( (\hat{u}, \hat{v}) \). For \( k = 1 \) these matrices are rows and represent the transposed gradient and the partial transposed gradient w.r.t. \( u \) of \( \vartheta \) at \( (\hat{u}, \hat{v}) \). Assume that \( k = 1 \) holds true, while \( \vartheta \) is twice differentiable. Then the matrices \( \nabla^2 \vartheta(\hat{u}, \hat{v}) \in \mathbb{R}^{(n+m) \times (n+m)} \), \( \nabla^2_u \vartheta(\hat{u}, \hat{v}) \in \mathbb{R}^{n \times n} \), and \( \nabla^2_{uv} \vartheta(\hat{u}, \hat{v}) \in \mathbb{R}^{n \times m} \) denote the Hessian, the Hessian w.r.t. \( u \) and the mixed Hessian w.r.t. \( u \) and \( v \) at \( (\hat{u}, \hat{v}) \), respectively. One may identify these matrices with the Fréchet derivatives introduced earlier.

Recall that a point \( \hat{u} \in U \) is called a Karush-Kuhn-Tucker-point, KKT-point for short, of the optimization problem

\[ \min \{ h(u) \mid H(u) \in C \}, \tag{3} \]

where \( h: U \longrightarrow \mathbb{R}, H: U \longrightarrow V \) are continuously Fréchet differentiable and \( C \subseteq V \) is a nonempty, closed, and convex cone, and set conv, provided it satisfies \( H(\hat{u}) \in C \) and there exists a multiplier \( \omega \in T_C(H(\hat{u}))^o \) such that \( h'(\hat{u}) + H'(\hat{u})^* \omega = 0 \). The latter conditions are called KKT-conditions of (3). Furthermore, the Kuczyłko-Robinson-Zowe-Constraint-Qualification, KRZCQ for short, is said to be valid at a feasible point \( u \in U \) of (3) whenever

\[ H'(u)[U] - R_C(H(u)) = V \]

is satisfied. The following results and their proofs are well-known in the theory of optimization in Banach spaces, see [31, Theorems 3.1 and 4.1] or [5, Theorem 3.9 and Proposition 3.16].

Lemma 2.3 Let \( \hat{u} \in U \) be a feasible point of (3) where KRZCQ is satisfied.

1. If \( \hat{u} \) is a local optimal solution of (3), then \( \hat{u} \) is a KKT-point of (3).

2. The nonempty, closed, convex cone

\[ A_0(\hat{u}) := \{ \omega \in T_C(H(\hat{u}))^o \mid H'(\hat{u})^* \omega = 0 \}, \]

which is called the set of singular Lagrange multipliers of (3) at \( \hat{u} \), contains only the zero vector \( 0 \).
3 MPCCs in Banach spaces

3.1 Formulation of the problem and failure of KRZCQ

In this section, we are going to study the complementarity constrained optimization problem

\[
\begin{align*}
a(x) \to & \, \min \\
b(x) \in & \, K \\
c(x) \in & \, C \\
d(x) \in & \, C^o \\
\langle c(x), d(x) \rangle & = 0.
\end{align*}
\]  

(4)

Therein, \(a : X \to \mathbb{R}, b : X \to W, c : X \to Z,\) and \(d : X \to Z^*\) are continuously Fréchet differentiable mappings between reflexive Banach spaces \(X, W,\) and \(Z.\) The set \(K \subseteq W\) is nonempty, closed, and convex, while \(C \subseteq Z\) is a nonempty, closed, convex cone. For an arbitrary point \(\hat{x} \in X\) which satisfies \(c(\hat{x}) \in C\) and \(d(\hat{x}) \in C^o,\) we already have \(\langle c(\hat{x}), d(\hat{x}) \rangle \leq 0.\) Hence, the last three constraints in (4) are called complementarity constraint. That is why (4) is often referred to as mathematical program with complementarity constraint, abbreviated by MPCC.

Let us define the set \(M \subseteq X\) which contains all feasible points of the MPCC (4) by means of

\[
M := \{x \in X \mid b(x) \in K, c(x) \in C, d(x) \in C^o, \langle c(x), d(x) \rangle = 0\}.
\]

The next lemma yields that KRZCQ fails to hold at any point of \(M.

Lemma 3.1 The constraint qualification KRZCQ is violated at any feasible point \(\hat{x} \in M\) of (4).

Proof We proceed by showing that there exist nonzero singular Lagrange multipliers at \(\hat{x} \in M\). By means of Lemma 2.3, this shows the failure of KRZCQ.

The quadruple \((\varrho, \nu, \mu, \xi) \in T_K(b(\hat{x}))^o \times C^o \times C \times \mathbb{R}\) is a singular Lagrange multiplier of \(\hat{x} \in M\) provided

\[
\varrho_X := b'(\hat{x})^* [\varrho] + c'(\hat{x})^* [\nu + \xi \cdot d(\hat{x})] + \langle c'(\hat{x})^* [d(\hat{x})] + d'(\hat{x})^* [c(\hat{x})], \rangle
\]

is satisfied. The above relation can be transformed into

\[
0 = \langle c(\hat{x}), \nu \rangle = \langle \mu, d(\hat{x}) \rangle = 0.
\]

Observe that \(\varpi := \varrho_{W^*}, \varpi := d(\hat{x}), \pi := c(\hat{x}),\) and \(\xi := -1\) satisfy all these conditions due to the complementarity condition \(\langle c(\hat{x}), d(\hat{x}) \rangle = 0.\) Hence, \((\varrho, \nu, \mu, \xi)\) is a nonvanishing singular Lagrange multiplier of (4), i.e. KRZCQ fails to hold at \(\hat{x} \in M\) according to Lemma 2.3.  \(\square

3.2 Auxiliary problems and stationarity concepts

In this section, we present the definitions of weak and strong stationarity for the MPCC (4) as defined in [26, Section 5.1]. Moreover, we give some constraint qualifications (CQs) which imply that these stationarity concepts are necessary optimality conditions. To this end, we follow [26] and introduce auxiliary problems without a complementarity constraint, which can be used in order to define the stationarity concepts for problem (4). The definition of these auxiliary problems parallels the corresponding definition in the finite-dimensional situation, see [26, Section 5.1].

Let \(\hat{x} \in M\) be an arbitrary feasible point of (4). We define the relaxed nonlinear problem, abbreviated by RNLP, of (4) at \(\hat{x}\) by

\[
\begin{align*}
a(x) \to & \, \min \\
b(x) \in & \, K \\
c(x) \in & \, C \cap d(\hat{x})^\perp \\
d(x) \in & \, C^o \cap c(\hat{x})^\perp.
\end{align*}
\]  

(5)
Obviously, \( \hat{x} \) is a feasible point of (5) but an arbitrary feasible point \( x \in X \) of the latter problem may not satisfy the complementarity condition \( \langle c(x), d(x) \rangle = 0 \). On the other hand, a feasible point of (4) is, in general, not feasible for (5).

In order to guarantee that the complementarity constraint \( \langle c(x), d(x) \rangle = 0 \) holds, we have to modify (5) by shrinking its feasible set. Therefore, we introduce the tightened nonlinear problem, TNLP for short, of (4) at \( \hat{x} \) by

\[
\begin{align*}
a(x) & \to \min \\
b(x) & \in K \\
c(x) & \in C \cap d(\hat{x})^\perp \cap (C^o \cap c(\hat{x})^\perp)^\perp \\
d(x) & \in C^o \cap c(\hat{x})^\perp \cap (C \cap d(\hat{x})^\perp)^\perp.
\end{align*}
\]

The feasible set of (6) is a subset of the feasible set \( M \) of (4) and contains \( \hat{x} \). Hence, if \( \hat{x} \in M \) is a local minimizer of (4), it is a local minimizer for (6). Consequently, under suitable regularity conditions it is a KKT-point of (6). In order to characterize KKT-points of (6), we need to calculate the polar cones of

\[
C \cap d(\hat{x})^\perp \cap (C^o \cap c(\hat{x})^\perp)^\perp \quad \text{and} \quad C^o \cap c(\hat{x})^\perp \cap (C \cap d(\hat{x})^\perp)^\perp.
\]

We summarize the corresponding calculations in the next lemma.

**Lemma 3.2** Let \( \hat{x} \in M \) be a feasible point of the MPCC (4). Then

\[
\begin{align*}
\text{lin}(d(\hat{x})) & \subseteq \text{lin}(C^o \cap c(\hat{x})^\perp) & \text{(7a)} \\
\text{lin}(c(\hat{x})) & \subseteq \text{lin}(C \cap d(\hat{x})^\perp) & \text{(7b)} \\
(C \cap d(\hat{x})^\perp \cap (C^o \cap c(\hat{x})^\perp)^\perp)^\circ & = \text{cl}(C^o - C^o \cap c(\hat{x})^\perp) & \text{(7c)} \\
(C^o \cap c(\hat{x})^\perp \cap (C \cap d(\hat{x})^\perp)^\perp)^\circ & = \text{cl}(C - C \cap d(\hat{x})^\perp). & \text{(7d)}
\end{align*}
\]

**Proof** From the feasibility of \( \hat{x} \in X \) we deduce \( d(\hat{x}) \in C^o \) and \( \langle c(\hat{x}), d(\hat{x}) \rangle = 0 \), i.e. \( d(\hat{x}) \in c(\hat{x})^\perp \). This leads to \( d(\hat{x}) \in C^o \cap c(\hat{x})^\perp \). (7a) follows from the monotonicity of the operator \( \text{lin}(\cdot) \). Analogously, we obtain (7b).

Let us show (7c). Observe that due to the inclusion \( \text{lin}(d(\hat{x})) \subseteq \text{lin}(C^o \cap c(\hat{x})^\perp) \), the equality

\[
\text{lin}(d(\hat{x})) + \text{lin}(C^o \cap c(\hat{x})^\perp) = \text{lin}(C^o \cap c(\hat{x})^\perp)
\]

holds. Due to the property of \( C^o \) to be a convex cone, we have \( C^o + C^o \cap c(\hat{x})^\perp = C^o \). Finally, recall that for any convex cone \( C \) the relation \( \text{lin}(C) = C - C \) is satisfied. Putting these three observations together, we obtain

\[
\begin{align*}
(C \cap d(\hat{x})^\perp \cap (C^o \cap c(\hat{x})^\perp)^\perp)^\circ & = \text{cl}(C^o + \text{lin}(d(\hat{x})) + \text{lin}(C^o \cap c(\hat{x})^\perp)) \\
& = \text{cl}(C^o + \text{cl}(C^o \cap c(\hat{x})^\perp)) \\
& = \text{cl}(C^o + C^o \cap c(\hat{x})^\perp - C^o \cap c(\hat{x})^\perp) \\
& = \text{cl}(C^o - C^o \cap c(\hat{x})^\perp).
\end{align*}
\]

Statement (7d) follows similarly. 

According to [26, Section 5.1], we define the weak and strong stationarity conditions for (4) by the KKT-conditions of the TNLP (6) and of the RNLP (5), respectively.

**Definition 3.1** Let \( \hat{x} \in M \) be an arbitrary feasible point of the MPCC (4).

1. The point \( \hat{x} \) is called **weakly stationary** provided there exist Lagrange multipliers \( \varrho \in T_K(b(\hat{x})^\circ) \), \( \nu \in \text{cl}(C^o - C^o \cap c(\hat{x})^\perp) \), and \( \mu \in \text{cl}(C - C \cap d(\hat{x})^\perp) \) which solve the system

\[
\begin{align*}
0 = a'(\hat{x}) + b'(\hat{x})^\ast[\varrho] + c'(\hat{x})^\ast[\nu] + d'(\hat{x})^\ast[\mu] \\
0 = \langle c(\hat{x}), \nu \rangle \\
0 = \langle \mu, d(\hat{x}) \rangle.
\end{align*}
\]

2. The point \( \hat{x} \) is called **strongly stationary** provided there exist Lagrange multipliers \( \varrho \in T_K(b(\hat{x})^\circ) \), \( \nu \in T_{C^o}(d(\hat{x})) \), and \( \mu \in T_C(c(\hat{x})) \) which solve system (8).
From the theory of finite-dimensional MPCCs we expect any strongly stationary point of (4) to be weakly stationary as well. The following lemma shows that this conjecture is correct.

**Lemma 3.3** If $\hat{x} \in M$ is a strongly stationary point of (4), then it is a weakly stationary point of (4) as well.

**Proof** Since $d(\hat{x}) \in C^o \cap c(\hat{x})^\perp$ holds, we have $C^o + \text{lin}(d(\hat{x})) \subseteq C^o - C^o \cap c(\hat{x})^\perp$ as well. Consequently,

$$\mathcal{T}_{C^o}(d(\hat{x})) = \text{cl}(C^o + \text{lin}(d(\hat{x}))) \subseteq \text{cl}(C^o - C^o \cap c(\hat{x})^\perp)$$

is satisfied. Similarly, we can prove $\mathcal{T}_C(c(\hat{x})) = \text{cl}(C + \text{lin}(c(\hat{x}))) \subseteq \text{cl}(C - C \cap d(\hat{x})^\perp)$. This shows the claim. $\Box$

It has been presented in [26] that the above notion of strong stationarity is an appropriate generalization of the finite-dimensional concept in the case where the cone $C$ is polyhedral. The next lemma shows that the definition of weak stationarity generalizes its finite-dimensional counterpart in a convincing way. We refer the reader to [25, 29] for several stationarity concepts (besides strong and weak stationarity) for finite-dimensional MPCCs. Furthermore, in [29] the author gives an overview of existing constraint qualifications for such problems.

**Lemma 3.4** Let $X = \mathbb{R}^n$, $W = \mathbb{R}^m$, $Z = \mathbb{R}^k$, $K = -\mathbb{R}_0^{m_1} \times \{0\}$, and $C = \mathbb{R}_0^{k_1}$ be given. Here, $m = m_1 + m_2$ shall hold for $m_1, m_2 \in \mathbb{N}$. We choose a feasible point $\hat{x} \in M$ of the corresponding MPCC (4). Let $c_1, \ldots, c_k, d_1, \ldots, d_k : \mathbb{R}^n \rightarrow \mathbb{R}$ denote the component functions of $c$ and $d$, respectively. Moreover, we introduce the index sets

$$I^{+0}(\hat{x}) := \{i \in \{1, \ldots, k\} \mid c_i(\hat{x}) > 0 \land d_i(\hat{x}) = 0\}$$

$$I^{00}(\hat{x}) := \{i \in \{1, \ldots, k\} \mid c_i(\hat{x}) = 0 \land d_i(\hat{x}) = 0\}$$

$$I^{0+}(\hat{x}) := \{i \in \{1, \ldots, k\} \mid c_i(\hat{x}) = 0 \land d_i(\hat{x}) < 0\}.$$

The point $\hat{x}$ is weakly stationary in the sense of Definition 3.1 if and only if there exist multipliers $\rho \in \mathbb{R}^m$ and $\nu, \mu \in \mathbb{R}^k$ which solve the system

$$0 = \nabla a(\hat{x})^\top + \nabla b(\hat{x})^\top \rho + \nabla c(\hat{x})^\top \nu + \nabla d(\hat{x})^\top \mu$$

$$\forall l \in \{1, \ldots, m_1\} : \quad g_l \geq 0$$

$$b(\hat{x})^\top \rho = 0$$

$$\forall i \in I^{+0}(\hat{x}) : \quad \nu_i = 0$$

$$\forall i \in I^{0+}(\hat{x}) : \quad \mu_i = 0. \quad (9)$$

**Proof** Due to the definition of weak stationarity and $(\mathbb{R}_0^{k_1})^\circ = -\mathbb{R}_0^{k_1}$, we only have to verify

$$\left(\mathbb{R}_0^{k_1} \cap c(\hat{x})^\perp - \mathbb{R}_0^{k_1}\right) \cap c(\hat{x})^\perp = \{\nu \in \mathbb{R}^k \mid \forall i \in I^{+0}(\hat{x}) : \nu_i = 0\}$$

$$\left(\mathbb{R}_0^{k_1} - \mathbb{R}_0^{k_1} \cap d(\hat{x})^\perp\right) \cap d(\hat{x})^\perp = \{\mu \in \mathbb{R}^k \mid \forall i \in I^{0+}(\hat{x}) : \mu_i = 0\}$$

(the closedness of the sets on the left-hand side is obvious). We just show the first equation, the second one can be handled analogously. This first equality, however, follows from

$$\left(\mathbb{R}_0^{k_1} \cap c(\hat{x})^\perp - \mathbb{R}_0^{k_1}\right) \cap c(\hat{x})^\perp = \mathbb{R}_0^{k_1} \cap c(\hat{x})^\perp - \mathbb{R}_0^{k_1} \cap c(\hat{x})^\perp$$

and

$$\mathbb{R}_0^{k_1} \cap c(\hat{x})^\perp = \{\nu \in \mathbb{R}_0^{k_1} \mid \forall i \in I^{+0}(\hat{x}) : \nu_i = 0\}.$$
Theorem 3.1 Let \( \hat{x} \in M \) be a local minimizer of the MPCC (4).

1. Let the convex cone \( C \subseteq W \times Z \times Z^* \) be defined by

\[
C := \mathcal{R}_K(b(\hat{x})) \times \left( \mathcal{R}_C(c(\hat{x})) \cap (-\mathcal{K}_C(c(\hat{x}), d(\hat{x}))) \right)
\times \left( \mathcal{R}_C^-(d(\hat{x})) \cap (-\mathcal{K}_C^-(d(\hat{x}), c(\hat{x}))) \right),
\]

and assume that the following qualification condition holds:

\[
H'(\hat{x})|X] - C = W \times Z \times Z^*.
\] (10)

Then \( \hat{x} \) is a weakly stationary point of (4).

2. Assume that \( H'(\hat{x}) \) is a surjective operator. Then \( \hat{x} \) is a strongly stationary point of (4).

Proof We show that condition (10) equals KRZCQ for (6). Due to the fact that \( \hat{x} \) is a local minimizer of (6) as well and the weak stationarity conditions equal the KKT-conditions for that problem, Lemma 2.3 yields the proof for the first statement.

Observe that if we define cones \( C_1 \subseteq Z \) and \( C_2 \subseteq Z^* \) by

\[
C_1 := C \cap d(\hat{x})^\perp \cap (C^\circ \cap c(\hat{x})^\perp)^\perp + \text{lin}(c(\hat{x}))
\]

\[
C_2 := C^\circ \cap c(\hat{x})^\perp \cap (C \cap d(\hat{x})^\perp)^\perp,
\]

then KRZCQ for (6) takes the form

\[
H'(\hat{x})|X] - \mathcal{R}_K(b(\hat{x})) \times \mathcal{R}_{C_1}(c(\hat{x})) \times \mathcal{R}_{C_2}(d(\hat{x})) = W \times Z \times Z^*.
\]

Hence, we just have to show

\[
\mathcal{R}_{C_1}(c(\hat{x})) = \mathcal{R}_C(c(\hat{x})) \cap (-\mathcal{K}_C(c(\hat{x}), d(\hat{x})))
\]

and

\[
\mathcal{R}_{C_2}(d(\hat{x})) = \mathcal{R}_C^-(d(\hat{x})) \cap (-\mathcal{K}_C^-(d(\hat{x}), c(\hat{x}))).
\]

We only verify the first of these equations since the proof of the second one is similar.

By (2d) and Lemma 2.2 we easily obtain

\[
\mathcal{R}_{C_1}(c(\hat{x})) = (C \cap d(\hat{x})^\perp \cap (C^\circ \cap c(\hat{x})^\perp)^\perp + \text{lin}(c(\hat{x}))
\]

\[
= (C + \text{lin}(c(\hat{x}))) \cap d(\hat{x})^\perp \cap (C^\circ \cap c(\hat{x})^\perp)^\perp
\]

\[
= (C + \text{lin}(c(\hat{x}))) \cap (\mathcal{T}_C(c(\hat{x}))^\perp)^\perp
\]

\[
= \mathcal{R}_C(c(\hat{x})) \cap d(\hat{x})^\perp \cap \mathcal{T}_C(c(\hat{x})) \cap (-\mathcal{T}_C(c(\hat{x})))
\]

\[
= \mathcal{R}_C(c(\hat{x})) \cap (-\mathcal{K}_C(c(\hat{x}), d(\hat{x}))).
\]

Hence, the proof of the first statement is completed.

The second statement is exactly [26, Proposition 5.2]. \( \square \)

Remark 3.1 Consider a feasible point \( \hat{x} \in M \) of the MPCC (4) in finite dimensions as introduced in Lemma 3.4. Then the constraint qualification (10) equals MPCC-MFCQ, i.e. for any \( g \in \mathbb{R}^m \) and any \( \nu, \mu \in \mathbb{R}^k \) we have

\[
0 = \nabla b(\hat{x})^\top g + \nabla c(\hat{x})^\top \nu + \nabla d(\hat{x})^\top \mu
\]

\[
\forall i \in \{1, \ldots, m_1\}: \quad g_i \geq 0
\]

\[
b(\hat{x})^\top g = 0
\]

\[
\forall i \in I^{+0}(\hat{x}): \quad \nu_i = 0
\]

\[
\forall i \in I^{0+}(\hat{x}): \quad \mu_i = 0.
\]

As mentioned in [13], MPCC-MFCQ implies a local optimal solution \( \hat{x} \) of (4) to be M-stationary, i.e. there exist \( g \in \mathbb{R}^m \) and \( \nu, \mu \in \mathbb{R}^k \) which satisfy the conditions in (9) as well as

\[
\forall i \in I^{00}(\hat{x}): \quad (\nu_i < 0 \land \mu_i > 0) \lor \nu_i \cdot \mu_i = 0.
\]
The latter stationarity condition is stronger than weak stationarity (cf. [13]). However, it is not clear how to define the concept of M-stationarity for generalized MPCCs.

The surjectivity of $H'(\hat{x})$ is a sufficient condition for MPCC-LICQ, i.e. the linear independence of the set

$$\{\nabla b_l(\hat{x}) \mid l \in J(\hat{x})\} \cup \{\nabla c_l(\hat{x}) \mid i \in I^0(\hat{x}) \cup I^0+(\hat{x})\} \cup \{\nabla d_l(\hat{x}) \mid i \in I^{00}(\hat{x}) \cup I^{00}(\hat{x})\}. $$

Therein, $b_1, \ldots, b_m: \mathbb{R}^n \rightarrow \mathbb{R}$ denote the $m$ component functions of $b$, while the active index set of $b$ at $\hat{x}$ is defined by means of $J(\hat{x}) := \{l \in \{1, \ldots, m\} \mid b_l(\hat{x}) = 0\}$. Note that MPCC-LICQ possesses the equivalent representation

$$0 = \nabla b(\hat{x})^T g + \nabla c(\hat{x})^T ν + \nabla d(\hat{x})^T μ \quad \forall l \notin J(\hat{x}): \quad g_l = 0$$
$$\forall i \in I^{0}(\hat{x}): \quad ν_i = 0$$
$$\forall i \in I^{0+}(\hat{x}): \quad μ_i = 0 \quad \implies \quad \begin{cases} g = 0 \\ ν = 0 \\ μ = 0. \end{cases}$$

According to [28], MPCC-LICQ is still a sufficient condition for a local optimal solution $\hat{x}$ of (4) to be strongly stationary, and we remark that this also follows from [26, Theorem 5.1].

Remark 3.2 From the theory of finite-dimensional MPCCs one could expect that strong stationarity and the KKT-conditions of (4) are equivalent. Unfortunately, it has been presented in [26] that strong stationarity is in general a weaker condition than the KKT-conditions of (4) and without a polyhedricity assumption on the cone $C$, it may happen that the strong stationarity conditions are too weak to yield a good necessary optimality condition (cf. [26, Section 5]). On the other hand, a linearization of the original problem may improve the situation as presented by means of two examples in [26, Section 6].

Although weak stationarity is a weaker condition than strong stationarity (cf. Lemma 3.3) and, hence, weaker than the KKT-conditions of the original MPCC (cf. Remark 3.2), it is still useful in order to find possible minimizer candidates of the problem (4). Note that the qualification condition presented in (10) is weaker than the surjectivity of the operator $H'(\hat{x})$ which was required in order to guarantee that a local minimizer $\hat{x} \in M$ of (4) is strongly stationary.

4 Stationarity conditions for bilevel programming problems

In this section, we study the bilevel programming problem

$$F(x, y) \rightarrow \min_{x, y} \quad G(x) \in K_u \quad y \in \Psi(x) \quad (11)$$

where $\Psi: X \rightarrow 2^Y$ is the solution set mapping of the parametric optimization problem

$$f(x, y) \rightarrow \min_y \quad g(x, y) \in K_l. \quad (12)$$

Therein, we assume that the Banach spaces $X$, $Y$, $W$, and $Z$ are reflexive, while $K_u \subseteq W$ and $K_l \subseteq Z$ are nonempty, closed, and convex cones. Moreover, we postulate $F: X \times Y \rightarrow \mathbb{R}$ and $G: X \rightarrow W$ to be continuously Fréchet differentiable while $f: X \times Y \rightarrow \mathbb{R}$ and $g: X \times Y \rightarrow Z$ have to be twice continuously Fréchet differentiable. At any point $(x, y) \in X \times Y$ which is feasible for the lower level problem (12), i.e., which satisfies $g(x, y) \in K_l$, we assume that KRZCQ for the lower level shall hold:

$$g'_b(x, y)[Y] - \mathcal{R}_{K_l}(g(x, y)) = Z.$$

Finally, for any $x \in X$, the mapping $f(x, \cdot): Y \rightarrow \mathbb{R}$ is assumed to be convex, while $g(x, \cdot): Y \rightarrow Z$ shall be $-K_l$-convex [18, Definition 2.4], i.e. for any $y, y' \in Y$ and any $\sigma \in [0, 1]$, we claim

$$g(x, σy + (1 − σ)y′) − σ \cdot g(x, y) − (1 − σ) \cdot g(x, y′) \in K_l.$$
The above assumptions guarantee that the KKT-conditions of (12) are necessary and sufficient for (global) optimality, that is

\[ y \in \Psi(x) \iff \exists \lambda \in K^0 \cap \text{int}K^0 : f'_y(x, y) + g'_y(x, y)^*[\lambda] = o_{y^*} \land g(x, y) \in K_i \land \langle g(x, y), \lambda \rangle = 0. \]

As a result, it is a common approach to replace problem (11) by its KKT-reformulation

\[
F(x, y) \to \min_{x, y, \lambda} \\
G(x) \in K_u \\
f'_y(x, y) + g'_y(x, y)^*[\lambda] = o_{y^*} \\
g(x, y) \in K_i \\
\lambda \in K^0 \\
\langle g(x, y), \lambda \rangle = 0,
\]

which is an MPCC as presented in Section 3. It has been mentioned in [10] that (11) and (13) are equivalent w.r.t. global optimal solutions (in a certain sense) and that any local optimal solution of (11) corresponds to a local optimal solution of (13). The converse of the last statement is not true in general but can be verified under additional assumptions in finite-dimensional spaces. However, the corresponding proof heavily relies on upper semicontinuity properties of the Lagrange multiplier mapping in the infinite-dimensional situation. Furthermore, one may check [9] for a detailed introduction to the theory of bilevel programming.

We refer to [5, Lemma 4.44 and Proposition 4.47] for results which ensure the upper semicontinuity of the Lagrange multiplier mapping in infinite-dimensional spaces. We refer to [5, Lemma 4.44 and Proposition 4.47] for results which ensure the upper semicontinuity of the Lagrange multiplier mapping in the infinite-dimensional situation. Furthermore, one may check [9] for a detailed introduction to the theory of bilelevel programming.

We adapt Definition 3.1 to problem (13) in order to introduce stationarity concepts for the bilevel programming problem (11).

**Definition 4.1** Let \((\hat{x}, \hat{y}) \in X \times Y\) be a feasible point of the bilevel programming problem (11).

1. The point \((\hat{x}, \hat{y})\) is called weakly stationary for (11) provided there exist multipliers \(\alpha \in K^0, \beta \in Y, \lambda \in K^0, \nu \in \text{cl}(K^0 \cap g(\hat{x}, \hat{y})^\perp), \) and \(\mu \in \text{cl}(K_i - K_i \cap \lambda^\perp)\) which solve the system

\[
\begin{align*}
o_{x^*} &= F'_x(\hat{x}, \hat{y}) + G'(\hat{x}, \hat{y})^*[\alpha] + f^{(2)}(\hat{x}, \hat{y})^*[\beta] + g'_x(\hat{x}, \hat{y})^*[\lambda] + g'_y(\hat{x}, \hat{y})^*[\nu] \\
o_{y^*} &= F'_y(\hat{x}, \hat{y}) + f^{(2)}(\hat{x}, \hat{y})^*[\beta] + g'_y(\hat{x}, \hat{y})^*[\lambda] + g'_y(\hat{x}, \hat{y})^*[\nu] \\
o_{z^*} &= g'_y(\hat{x}, \hat{y})^*[\beta] + \mu \\
0 &= \langle G(\hat{x}), \alpha \rangle = \langle g(\hat{x}, \hat{y}), \lambda \rangle = \langle g(\hat{x}, \hat{y}), \nu \rangle = \langle \mu, \lambda \rangle.
\end{align*}
\]

2. The point \((\hat{x}, \hat{y})\) is called strongly stationary for (11) provided there exist multipliers \(\alpha \in K^0, \beta \in Y, \lambda \in K^0, \nu \in \mathcal{T}_{K_i}(\lambda), \) and \(\mu \in \mathcal{T}_{K_i}(g(\hat{x}, \hat{y}))\) which satisfy the conditions in (14).

Observe that a feasible point \((\hat{x}, \hat{y}) \in X \times Y\) of (11) is weakly (resp. strongly) stationary for (11) w.r.t. Definition 4.1 if and only if there is a lower level multiplier \(\lambda \in Z^\ast\) such that \((\hat{x}, \hat{y}, \lambda)\) is feasible for (13) and a weakly (resp. strongly) stationary point of that problem w.r.t. Definition 3.1. It is not difficult to check that the above stationarity concepts generalize the corresponding conditions in finite dimensions stated, e.g., in [11]. We omit the proofs here since they follow from Lemma 3.4 and [26]. Especially, it is possible to eliminate the multiplier \(\mu\) from the above stationarity definition in finite dimensions. The same is possible in arbitrary reflexive Banach spaces. The corresponding proofs are straightforward and, hence, omitted as well.

**Remark 4.1** Let \((\hat{x}, \hat{y}) \in X \times Y\) be a feasible point of the bilevel programming problem (11).
1. The point \((\bar{x}, \bar{y})\) is weakly stationary for (11) if and only if there exist multipliers \(\alpha \in K^\circ_\ell \), \(\beta \in Y\), \(\lambda \in K^\circ_l\), and \(\nu \in \text{cl}(K^\circ_l - K^\circ_l \cap g(\bar{x}, \bar{y})^-)\) which satisfy \(-g^\alpha_y(\bar{x}, \bar{y})[\beta] \in \text{cl}(K_\ell - K_\ell \cap \lambda^-)\) as well as the following conditions:

\[
\begin{align*}
0 &= F^x(\bar{x}, \bar{y}) + G^x(\bar{x}, \bar{y})[\alpha] + f^x_{g^x}(\bar{x}, \bar{y})[\beta] \\
&\quad + (g^x_{g^x}(\bar{x}, \bar{y})[\beta])^\top[\lambda] + g^x_\nu(\bar{x}, \bar{y})^\top[\nu] \\
0 &= F^y(\bar{x}, \bar{y}) + f^y_\nu(\bar{x}, \bar{y})[\beta] + (g^y_{g^y}(\bar{x}, \bar{y})[\beta])^\top[\lambda] + g^y_\nu(\bar{x}, \bar{y})^\top[\nu] \\
0 &= \langle G(\bar{x}), \alpha \rangle = \langle g(\bar{x}, \bar{y}), \lambda \rangle = \langle g(\bar{x}, \bar{y}), \nu \rangle = \langle g^\alpha_y(\bar{x}, \bar{y})[\beta], \lambda \rangle.
\end{align*}
\] (15)

2. The point \((\bar{x}, \bar{y}) \in X \times Y\) is strongly stationary for (11) if and only if there exist multipliers \(\alpha \in K^\circ_\ell\), \(\beta \in Y\), \(\lambda \in K^\circ_l\), and \(\nu \in T_{K^\circ_l}(\lambda)\) which satisfy \(-g^\alpha_y(\bar{x}, \bar{y})[\beta] \in T_{K_\ell}(g(\bar{x}, \bar{y}))\) and the conditions in (15).

The construction of constraint qualifications for (11), which ensure that a local optimal solution of the latter problem is weakly or strongly stationary, turns out to be difficult. Such qualification conditions can be expected to depend on the lower level multiplier \(\lambda\), which was used for the transformation into (13), since any stationarity condition for (11) is in fact a stationarity condition for the KKT-reformulation (13). Hence, most regularity conditions in literature on finite-dimensional bilevel programming problems just correspond to problem (13) and not to (11) directly.

However, if the lower level possesses constraints of special affine type, the situation is more comfortable.

**Proposition 4.1** Let \((\bar{x}, \bar{y}) \in X \times Y\) be a local optimal solution of the bilevel programming problem (11) where the mapping \(g: X \times Y \longrightarrow Z\) is given by

\[
\forall (x, y) \in X: \quad g(x, y) := g_1(x) + B[y].
\]

Here, \(g_1: X \longrightarrow Z\) is continuously Fréchet differentiable and \(B \in \mathbb{L}[Y, Z]\) is a bounded linear operator.

Assume that the linear operator \(Q \in \mathbb{L}[X \times Y, W \times Y^* \times Z]\) is defined by

\[
Q[d_x, d_y] := \left(G^x(\bar{x})[d_x], f^y_{g^y}(\bar{x}, \bar{y})[d_y], g^x_\nu(\bar{x}, \bar{y})[d_x] + B[d_y]\right)
\]

for all \((d_x, d_y) \in X \times Y\) is surjective. Then \((\bar{x}, \bar{y})\) is a strongly stationary point for the corresponding bilevel programming problem.

**Proof** First, we find some \(\lambda \in K^\circ_l\) such that \((\bar{x}, \bar{y}, \lambda)\) is a local optimal solution of the corresponding KKT-reformulation (13). Now we are going to apply the second statement of Theorem 3.1. Hence, we introduce \(H: X \times Y \times Z^* \longrightarrow W \times Y^* \times Z \times Z^*\) by means of

\[
H(x, y, z^*) := \langle G(x), f^y_\nu(x, y) + B^*[z^*], g_1(x) + B[y], z^* \rangle
\]

for any \((x, y, z^*) \in X \times Y \times Z^*\). Computing its Fréchet derivative at \((\bar{x}, \bar{y}, \lambda)\) yields

\[
H'(\bar{x}, \bar{y}, \lambda)[d_x, d_y, d_{z^*}] = \langle G'(\bar{x})[d_x], f^y_{g^y}(\bar{x}, \bar{y})[d_y], g_1'(\bar{x})[d_x] + B'[d_y], d_{z^*} \rangle
\]

for any \((d_x, d_y, d_{z^*}) \in X \times Y \times Z^*\). Observe that \(H'(\bar{x}, \bar{y}, \lambda)\) does not depend on \(\lambda\) in this special setting.

Take an arbitrary point \((w, y^*, z, z^*) \in W \times Y^* \times Z \times Z^*\), and consider the linear equation

\[
H'(\bar{x}, \bar{y}, \lambda)[d_x, d_y, d_{z^*}] = (w, y^*, z, z^*)
\]

Obviously, we have \(d_{z^*} = z^*\), and the above equation can be reduced to

\[
Q[d_x, d_y] = (w, y^* - B^*[z^*], z).
\]

The latter linear equation possesses a solution since \(Q\) is assumed to be surjective. Hence, \(H'(\bar{x}, \bar{y}, \lambda)\) is surjective and due to Theorem 3.1 \((\bar{x}, \bar{y}, \lambda)\) is strongly stationary for (13), which means by definition that \((\bar{x}, \bar{y})\) is strongly stationary for the bilevel programming problem (11). \(\square\)
Example 4.1 We fix $X = \mathbb{R}^n$, $Y = \mathbb{R}^m$, $W = \mathbb{R}^p$, $Z = \mathbb{R}^q$, $K_u = -\mathbb{R}^{p,+}$, and $K_l = -\mathbb{R}^{q,+}$ as well as
\[
G(x) := Cx - c \quad f(x, y) := \frac{1}{2}y^\top Ry + y^\top Px \quad g(x, y) := Ax + By - d
\]
for all $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$ and given matrices of appropriate dimension such that $R$ is a symmetric and positively semidefinite matrix. Assume that the matrix
\[
Q := \begin{bmatrix} C & O \\ P & R \\ A & B \end{bmatrix} \in \mathbb{R}^{(p+m+q) \times (n+m)}
\]
possesses full row rank $p + m + q$ (note that a necessary condition for that property is $p + q \leq n$). Then any local optimal solution $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^m$ of the corresponding bilevel programming problem is strongly stationary, i.e. there exist Lagrange multipliers $\alpha \in \mathbb{R}^{p,+}$, $\beta \in \mathbb{R}^m$, $\lambda \in \mathbb{R}^{q,+}$, and $\nu \in \mathbb{R}^q$ which satisfy the following set of conditions:
\[
\begin{align*}
0 &= \nabla_x F(\hat{x}, \hat{y})^\top + C^\top \alpha + P^\top \beta + A^\top \nu \\
0 &= \nabla_y F(\hat{x}, \hat{y})^\top + R \beta + B^\top \nu \\
0 &= R \hat{y} + P \hat{x} + B^\top \lambda \\
0 &= (C \hat{x} - c)^\top \alpha - (A \hat{x} + B \hat{y} - d)^\top \lambda \\
\forall i &\in I^{00}(\hat{x}, \hat{y}, \lambda): \; \nu_i = 0 \\
\forall i &\in I^{0+}(\hat{x}, \hat{y}, \lambda): \; (A \beta)_i = 0 \\
\forall i &\in I^{00}(\hat{x}, \hat{y}, \lambda): \; \nu_i \geq 0 \; \land \; (A \beta)_i \geq 0.
\end{align*}
\]
Therein, the appearing index sets are defined as stated below (cf. Lemma 3.4):
\[
\begin{align*}
I^{00}(\hat{x}, \hat{y}, \lambda) &:= \{ i \in \{1, \ldots, q\} \mid (A \hat{x} + B \hat{y} - d)_i < 0 \; \land \; \lambda_i = 0 \} \\
I^{00}(\hat{x}, \hat{y}, \lambda) &:= \{ i \in \{1, \ldots, q\} \mid (A \hat{x} + B \hat{y} - d)_i = 0 \; \land \; \lambda_i = 0 \} \\
I^{0+}(\hat{x}, \hat{y}, \lambda) &:= \{ i \in \{1, \ldots, q\} \mid (A \hat{x} + B \hat{y} - d)_i = 0 \; \land \; \lambda_i > 0 \}.
\end{align*}
\]
Note that the constraint qualification, i.e. the rank condition on $Q$, does not depend on the lower level multiplier $\lambda$. □

We generalize the above example to arbitrary reflexive Banach spaces.

Example 4.2 Choose operators $A \in \mathbb{L}[X, Z]$, $B \in \mathbb{L}[Y, Z]$, $C \in \mathbb{L}[X, W]$, a continuous, positive semidefinite, symmetric bilinear form $\mathcal{R} : Y \times Y \rightarrow \mathbb{R}$, a continuous bilinear form $\mathcal{P} : X \times Y \rightarrow \mathbb{R}$, and vectors $c \in W$ as well as $d \in Z$ such that
\[
G(x) := C[x] - c \quad f(x, y) := \frac{1}{2}\mathcal{R}[y, y] + \mathcal{P}[x, y] \quad g(x, y) = A[x] + B[y] - d
\]
holds true for all $(x, y) \in X \times Y$. If the linear operator $Q \in \mathbb{L}[X \times Y, W \times Y^* \times Z]$ defined by
\[
Q[d_x, d_y] := (C[d_x], \mathcal{P}[d_x, \cdot] + \mathcal{R}[d_y, \cdot], A[d_x] + B[d_y])
\]
for any $(d_x, d_y) \in X \times Y$ is surjective, then any local minimizer $(\hat{x}, \hat{y}) \in X \times Y$ of the corresponding bilevel programming problem is strongly stationary, i.e. there exist Lagrange multipliers $\alpha \in K_u^\circ$, $\beta \in Y$, $\lambda \in K_l^\circ$, and $\nu \in \mathcal{T}_{K_l}(\lambda)$ which satisfy $-B[\beta] \in \mathcal{T}_{K_u}(A[\hat{x}] + B[\hat{y}] - d)$ and the following set of conditions:
\[
\begin{align*}
o_{X^*} &= F'_x(\hat{x}, \hat{y}) + C^*[\alpha] + \mathcal{P}[\cdot, \beta] + A^*[\nu] \\
o_{Y^*} &= F'_y(\hat{x}, \hat{y}) + \mathcal{R}[\cdot, \beta] + B^*[\nu] \\
o_{Y^*} &= \mathcal{R}[\cdot, \beta] + \mathcal{P}[\hat{x}, \cdot] + B^*[\lambda] \\
0 &= \langle A[\hat{x}] + B[\hat{y}] - d, \lambda \rangle \\
0 &= \langle A[\hat{x}] + B[\hat{y}] - d, \nu \rangle \\
0 &= \langle B[\beta], \lambda \rangle.
\end{align*}
\]
Again we note that the constraint qualification, i.e. the surjectivity of $Q$, does not depend on the lower level multiplier $\lambda$. □
As mentioned above, constraint qualifications for general bilevel programming problems (11) ensuring local optimal solution to be weakly or strongly stationary are likely to depend on the lower level multiplier \( \lambda \). We underline this statement in the subsequent theorem.

**Theorem 4.1** Let \((\hat{x}, \hat{y}) \in X \times Y\) be a local optimal solution of the bilevel programming problem (11). Then the following statements hold.

1. Let \( \lambda \in \mathbb{Z}^* \) be chosen such that \((\hat{x}, \hat{y}, \lambda)\) is a feasible point of (13). Furthermore, suppose that the linear operator \( Q \in \mathbb{L}[X \times Y, W \times Y^* \times Z] \) defined by
   \[
   Q[d_x, d_y] := (G'(\hat{x})[d_x], f_{y^*}^{(2)}(\hat{x}, \hat{y})[d_x],] + (g_{y^*}^{(2)}(\hat{x}, \hat{y})[d_x,])^*[\lambda]
   \]
   for any \((d_x, d_y) \in X \times Y\) and the cone \( C \subseteq W \times Y^* \times Z \) given by
   \[
   C := \mathcal{R}_{K_i}(G(\hat{x})) \times \{\phi_{Y^*}\} \times \left(\mathcal{R}_{K_i}(g(\hat{x}, \hat{y})) \cap (-K_{K_i}(g(\hat{x}, \hat{y}), \lambda))\right)
   \]
   satisfy the constraint qualification
   \[
   Q[X \times Y] - C' = W \times Y^* \times Z.
   \]

   Then \((\hat{x}, \hat{y})\) is a weakly stationary point of (11).

2. Let \( \lambda \in \mathbb{Z}^* \) be chosen such that \((\hat{x}, \hat{y}, \lambda)\) is a feasible point of (13). Furthermore, suppose that the operator \( Q \in \mathbb{L}[X \times Y, W \times Y^* \times Z] \) defined within (17) is surjective. Then \((\hat{x}, \hat{y})\) is a strongly stationary point of (11).

**Proof** We start verifying the first statement. From the choice of \( \lambda \) and the definition of weak stationarity for the bilevel programming problem (11) we only have to show that \((\hat{x}, \hat{y}, \lambda)\) is a weakly stationary point for the MPCC (13). Therefore, we exploit the first statement of Theorem 3.1. In order to do so, we introduce a mapping \( H : X \times Y \times Z^* \rightarrow W \times Y^* \times Z^* \) by means of
\[
H(x, y, z^*) := (G(x), f'_y(x, y) + g'_y(x, y)^*[z^*], g(x, y), z^*)
\]
and a cone \( C \subseteq W \times Y^* \times Z^* \) as mentioned above:
\[
C := \mathcal{R}_{K_i}(G(\hat{x})) \times \{\phi_{Y^*}\} \times \left(\mathcal{R}_{K_i}(g(\hat{x}, \hat{y})) \cap (-K_{K_i}(g(\hat{x}, \hat{y}), \lambda))\right) \times \left(\mathcal{R}_{K_{l_i}}(\lambda) \cap (-K_{K_{l_i}}(\lambda, g(\hat{x}, \hat{y})))\right).
\]
From the aforementioned result it is clear that \((\hat{x}, \hat{y}, \lambda)\) is weakly stationary for (13) provided the constraint qualification
\[
H'(\hat{x}, \hat{y}, \lambda)[X \times Y \times Z^*] - C = W \times Y^* \times Z^*
\]
is satisfied. Hence, we take an arbitrary point \((w, y^*, z, z^*) \in W \times Y^* \times Z \times Z^*\) and show that there is a solution \((d_x, d_y, d_z^*) \in X \times Y \times Z^*\) of the following generalized equation:
\[
H'(\hat{x}, \hat{y}, \lambda)[d_x, d_y, d_z^*] - (w, y^*, z, z^*) \in C.
\]
Observe that
\[
H'(\hat{x}, \hat{y}, \lambda)[d_x, d_y, d_z^*] = (G'(\hat{x})[d_x], f_{y^*}^{(2)}(\hat{x}, \hat{y})[d_x,] + (g_{y^*}^{(2)}(\hat{x}, \hat{y})[d_x,])^*[\lambda]
\]
holds true. Fixing \(d_z^* = z^*\) and \(\phi_{Z^*} \in \mathcal{R}_{K_{l_i}}(\lambda) \cap (-K_{K_{l_i}}(\lambda, g(\hat{x}, \hat{y})))\), we can reduce (19) to
\[
Q[d_x, d_y] - (w, y^* - g'_y(\hat{x}, \hat{y})^*[z^*], z) \in C'.
\]
The latter equation possesses a solution from the assumption of this theorem. Consequently, \((\hat{x}, \hat{y}, \lambda)\) is weakly stationary for (13), i.e. \((\hat{x}, \hat{y})\) is weakly stationary for (11).

The proof of the second statement is similar to that one of the first assertion: we reduce the surjectivity of \(H'(\hat{x}, \hat{y}, \lambda)\) (which is a sufficient condition for strong stationarity by Theorem 3.1) to the surjectivity of \(Q\). □
Remark 4.2 Observe that the constraint qualification (18) postulated in Theorem 4.1 is stronger than (10) introduced for MPCCs in Theorem 3.1 but much easier to check since the image space has been reduced from $W \times Y^* \times Z \times Z^*$ to $W \times Y^* \times Z$. Within the proof of Theorem 4.1 one can find the operator $H'(\hat{x}, \hat{y}, \lambda)$ and the convex cone $\mathcal{C}$, which can be used to define the weaker constraint qualification

$$H'(\hat{x}, \hat{y}, \lambda)[X \times Y \times Z^*] - \mathcal{C} = W \times Y^* \times Z \times Z^*.$$  \hfill (20)

It is clear from the proof that (20) is sufficient for the weak stationarity of $(\hat{x}, \hat{y})$ as well.

Observe that the operator $Q$ from (17) is surjective if and only if $H'(\hat{x}, \hat{y}, \lambda)$ is surjective, i.e. the corresponding constraint qualifications, which imply strong stationarity of $(\hat{x}, \hat{y})$, are equivalent.

Remark 4.3 Recall that in (the linear) Example 4.2 the constraint qualification implying strong stationarity of any local optimal solution neither depends on the solution nor on a corresponding lower level multiplier. On the other hand, from Theorem 4.1 it is clear that weaker constraint qualifications of KRZCQ-type, which only imply weak stationarity of local optimal solutions, depend on the solution and a corresponding lower level multiplier.

Remark 4.4 Let $(\hat{x}, \hat{y}) \in X \times Y$ be a local optimal solution of (11) where the corresponding set of lower level multipliers is not a singleton. Then it may happen that the constraint qualifications postulated in Theorem 4.1 hold for some but not all multipliers, i.e. the choice of the multiplier $\lambda$ in this theorem is of essential importance.

Example 4.3 Let $X = Y = Z = \mathbb{R}^2$ and $K_l = -\mathbb{R}_{0}^{2:}$ hold, and consider the bilevel programming problem

$$\min_{x,y} \left\{ \|x\|_2^2 + \|y\|_2^2 \mid y \in \Psi(x) := \text{Argmin}\{ (y_2 - 1)^2 \mid y_2^2 + y_2 - x_1 \leq 0, y_2 - x_2 \leq 0 \} \right\}$$

whose lower level is convex and regular at any feasible point. Its unique global optimal solution is given by $(\hat{x}, \hat{y}) := (0, 0)$. One may check that the corresponding set of lower level multipliers is given by

$$\Lambda := \text{conv}\left\{ \{(0, 2)^{\top}, (2, 0)^{\top}\} \right\},$$

i.e. $(\hat{x}, \hat{y}, \lambda)$ is a global optimal solution of the corresponding KKT-reformulation for any $\lambda \in \Lambda$. Observe that the matrix $Q \in \mathbb{R}^{4 \times 4}$ corresponding to the operator $Q$ from (17) takes the form

$$Q = \begin{pmatrix} 0 & 0 & 2\lambda_1 & 0 \\ 0 & 0 & 0 & 2 \\ -1 & 0 & 0 & 1 \\ 0 & -1 & 0 & 1 \end{pmatrix}$$

and is singular for precisely one multiplier from $\Lambda$, namely $\hat{\lambda} = (0, 2)^{\top}$. Especially, both constraint qualifications introduced within Theorem 4.1 do not hold at $(\hat{x}, \hat{y}, \hat{\lambda})$ but at any other point $(\bar{x}, \bar{y}, \bar{\lambda})$ such that $\lambda \in \Lambda \setminus \{\hat{\lambda}\}$ is satisfied.

Remark 4.5 Let $X = \mathbb{R}^n$, $Y = \mathbb{R}^m$, $W = \mathbb{R}^q$, $Z = \mathbb{R}^q$, $K_u = -\mathbb{R}_{0}^{2:}$, and $K_l = -\mathbb{R}_{0}^{2:}$ hold, and choose an arbitrary feasible point $(\bar{x}, \bar{y}) \in \mathbb{R}^n \times \mathbb{R}^m$ of the corresponding bilevel programming problem (11).

It is weakly stationary in the sense of Definition 4.1 provided there exist multipliers $\alpha \in \mathbb{R}_{0}^{q:}, \beta \in \mathbb{R}^m, \lambda \in \mathbb{R}_{0}^{n:}$, and $\nu \in \mathbb{R}^q$ which solve the system

$$0 = \nabla_x F(\bar{x}, \bar{y})^{\top} + \nabla G(\bar{x})^{\top} \alpha + \nabla_{yx}^2 f(\bar{x}, \bar{y})^{\top} \beta$$

$$+ \sum_{i=1}^{q} \lambda_i \nabla_{yx} g_i(\bar{x}, \bar{y})^{\top} \beta + \nabla_x g(\bar{x}, \bar{y})^{\top} \nu,$$

$$0 = \nabla_y F(\bar{x}, \bar{y})^{\top} + \nabla_{yy}^2 f(\bar{x}, \bar{y})^{\top} \beta$$

$$+ \sum_{i=1}^{q} \lambda_i \nabla_{yy} g_i(\bar{x}, \bar{y})^{\top} \beta + \nabla_y g(\bar{x}, \bar{y})^{\top} \nu,$$

$$0 = \nabla_y f(\bar{x}, \bar{y})^{\top} + \nabla_{yy} g(\bar{x}, \bar{y})^{\top} \lambda,$$

$$G(\bar{x})^{\top} \alpha = 0 \quad g(\bar{x}, \bar{y})^{\top} \lambda = 0,$$

$\forall i \in I^{10}(\bar{x}, \bar{y}, \lambda): \quad \nu_i = 0,$

$\forall i \in I^{1+}(\bar{x}, \bar{y}, \lambda): \quad \nabla_y g_i(\bar{x}, \bar{y})^{\top} \beta = 0.$
Furthermore, \((\hat{x}, \hat{y})\) is strongly stationary in the sense of Definition 4.1 provided there exist multipliers \(\alpha \in \mathbb{R}_0^{n_1}, \beta \in \mathbb{R}^m, \lambda \in \mathbb{R}_0^{n_2},\) and \(\nu \in \mathbb{R}^q\) which satisfy (21) as well as
\[
\forall i \in I^{00}(\hat{x}, \hat{y}, \lambda) : \quad \nu_i \geq 0 \land \nabla_y g_i(\hat{x}, \hat{y}) \beta \geq 0.
\]
Let \(\lambda \in \mathbb{R}^q\) be chosen such that \((\hat{x}, \hat{y}, \lambda)\) is a feasible point for (13). Then for \(\alpha \in \mathbb{R}^p, \beta \in \mathbb{R}^m,\) and \(\nu \in \mathbb{R}^q\) the CQ (18) reads
\[
0 = \nabla G(\hat{x})^\top \alpha + \nabla_{yx}^2 f(\hat{x}, \hat{y})^\top \beta + \sum_{i=1}^q \nabla_{y}^2 g_i(\hat{x}, \hat{y})^\top \beta + \nabla_x g(\hat{x}, \hat{y})^\top \nu
\]
while (20) equals the corresponding MPCC-MFCQ condition, which takes the following form:
\[
0 = \nabla G(\hat{x})^\top \alpha + \nabla_{yx}^2 f(\hat{x}, \hat{y})^\top \beta + \sum_{i=1}^q \nabla_{y}^2 g_i(\hat{x}, \hat{y})^\top \beta + \nabla_x g(\hat{x}, \hat{y})^\top \nu
\]
Furthermore, the corresponding MPCC-LICQ is given below:
\[
0 = \nabla G(\hat{x})^\top \alpha + \nabla_{yx}^2 f(\hat{x}, \hat{y})^\top \beta + \sum_{i=1}^q \nabla_{y}^2 g_i(\hat{x}, \hat{y})^\top \beta + \nabla_x g(\hat{x}, \hat{y})^\top \nu
\]
Therein, the index sets \(I^{00}(\hat{x}, \hat{y}, \lambda), I^{00}(\hat{x}, \hat{y}, \lambda), I^{0+}(\hat{x}, \hat{y}, \lambda),\) and \(J(\hat{x})\) are defined by
\[
I^{00}(\hat{x}, \hat{y}, \lambda) := \{ i \in \{1, \ldots, q\} \mid g_i(\hat{x}, \hat{y}) < 0 \land \lambda_i = 0 \}
\]
\[
I^{00}(\hat{x}, \hat{y}, \lambda) := \{ i \in \{1, \ldots, q\} \mid g_i(\hat{x}, \hat{y}) = 0 \land \lambda_i = 0 \}
\]
\[
I^{0+}(\hat{x}, \hat{y}, \lambda) := \{ i \in \{1, \ldots, q\} \mid g_i(\hat{x}, \hat{y}) = 0 \land \lambda_i > 0 \}
\]
\[
J(\hat{x}) := \{ j \in \{1, \ldots, p\} \mid G_j(\hat{x}) = 0 \}
\]
where \(G_1, \ldots, G_p : \mathbb{R}^n \rightarrow \mathbb{R}\) and \(g_1, \ldots, g_q : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}\) denote the component functions of \(G\) and \(g\), respectively. Observe that (due to Remark 4.1) we eliminated the multiplier \(\mu\) from the regularity and stationarity conditions.

One may check [11, 30] for more information on stationarity concepts and constraint qualifications for finite-dimensional bilevel programming problems.
5 An application in bilevel optimal control

In this section, we want to derive necessary optimality conditions for a certain bilevel optimal control problem, which is given by means of

\[ F_0(x(T), y(T)) + \int_0^T F_1(t, x(t), y(t), u(t), v(t)) \, dt \rightarrow \min_{x, u, y, v} \]

\[ \dot{x}(t) - C_x x(t) - C_u u(t) = 0 \]

\[ x(0) - x_0 = 0 \]

\[ (y, v) \in \Psi(x, u) \]

where \( \Psi : W_{t,2}^n(0, T) \times L_2^k(0, T) \rightarrow 2^{W_{t,2}^n(0, T) \times L_2^l(0, T)} \) denotes the solution set mapping of the parametric lower level optimal control problem

\[ \frac{1}{2} y(T)^	op R_0 y(T) + \int_0^T \frac{1}{2} \left[ y(t)^	op \left[ R_y y(t) + 2P x(t) \right] + v(t)^	op \left[ R_v v(t) \right] \right] \, dt \rightarrow \min_{y, v} \]

\[ \dot{y}(t) - A_y x(t) - B_y y(t) - A_u u(t) - B_v v(t) = 0 \]

\[ y(0) - y_0 = 0 \]

\[ D_u u(t) + D_v v(t) - \delta(t) \leq 0. \]

Here and in what follows, we assume all the constraints appearing in (25), (26), and all derived surrogate problems to hold almost everywhere on \((0, T)\).

Nowadays, many practical problems turn out to have a hierarchical and dynamic structure, which is why their corresponding mathematical models are so-called bilevel optimal control problems. In general, a bilevel optimal control problem could be defined as a bilevel programming problem where at least one decision level is an optimal control problem. Hence, bilevel optimal control combines the difficulties of bilevel programming and infinite-dimensional optimization. There exist several publications introducing applications and numerical approaches to bilevel optimal control, cf. [1, 12, 14, 15, 20, 21, 23], but surprisingly few theoretical results and optimality conditions, which are not based on discretization, cf. [3, 4, 6, 7, 27]. Here we want to apply the theory of MPCCs in reflexive Banach spaces in order to derive necessary optimality conditions for the bilevel optimal control problem (25).

Below we list all the standing assumptions postulated on (25) and its lower level problem (26).

1. The function \( F_0 : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R} \) is continuously differentiable.
2. The function \( F_1 : (0, T) \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^k \times \mathbb{R}^l \rightarrow \mathbb{R} \) is continuously differentiable w.r.t. its last four components and measurable w.r.t. its first component. Furthermore, we assume the existence of constants \( c, c' > 0 \) such that for any \( t \in (0, T) \), \( x \in \mathbb{R}^n \), \( y \in \mathbb{R}^m \), \( u \in \mathbb{R}^k \), and \( v \in \mathbb{R}^l \) the following growth-conditions hold:

\[ |F_1(t, x, y, u, v)| \leq c \left( 1 + \|x\|^2 + \|y\|^2 + \|u\|^2 + \|v\|^2 \right) \]

\[ \|\nabla_{x,y,u,v} F_1(t, x, y, u, v)\|_2 \leq c' \left( 1 + \|x\|^2 + \|y\|^2 + \|u\|^2 + \|v\|^2 \right). \]

Consequently, \( F : W_{t,2}^n(0, T) \times L_2^k(0, T) \times W_{t,2}^m(0, T) \times L_2^l(0, T) \rightarrow \mathbb{R} \), the upper level objective function, is continuously Fréchet differentiable (cf. [8]).
3. The matrices \( A_x \in \mathbb{R}^{m \times n} \), \( A_u \in \mathbb{R}^{m \times k} \), \( B_y \in \mathbb{R}^{m \times m} \), \( B_u \in \mathbb{R}^{m \times l} \), \( C_x \in \mathbb{R}^{n \times n} \), \( C_u \in \mathbb{R}^{n \times k} \), \( D_u \in \mathbb{R}^{r \times k} \), \( D_v \in \mathbb{R}^{r \times l} \), \( P \in \mathbb{R}^{m \times m} \), \( R_0 \in \mathbb{R}^{m \times m} \), \( R_y \in \mathbb{R}^{m \times m} \), and \( R_v \in \mathbb{R}^{l \times l} \) are fixed. Furthermore, we assume \( R_0 \), \( R_y \) as well as \( R_v \) to be symmetric positive semidefinite and \( D_v \) to possess full row rank \( r \).
4. The function \( \delta \in L_2^l(0, T) \) as well as the initial states \( x_0 \in \mathbb{R}^n \) and \( y_0 \in \mathbb{R}^m \) are fixed.
5. We choose the state space \( W_{1,2}(0, T) \), i.e. the set of all absolutely continuous functions possessing a weak derivative which belongs to \( L_2(0, T) \). The latter set is the control space of the above bilevel optimal control problem.

Due to these assumptions, the lower level problem (26) is convex w.r.t. the lower level decision variables \( y \) and \( v \). It is not difficult to show that the overall problem (25) is of the type presented in Example 4.2. Using the abbreviations introduced in this example, the linear operator \( B \) is surjective,
Remark 5.1 Let \( (\hat{x}, \hat{u}) \in W_{1,2}^n(0,T) \times L_2^k(0,T) \) be an arbitrary point feasible for the upper level problem of (25). A pair of functions \( (\hat{y}, \hat{v}) \in W_{1,2}^m(0,T) \times L_2^l(0,T) \) is an optimal solution for (26) if and only if it is feasible for (26) and there exist functions \( \vartheta \in W_{1,2}^m(0,T) \) and \( \lambda \in L_2^l(0,T) \) such that the following conditions hold for almost every \( t \in (0,T) \):
\[
\begin{align*}
\dot{\vartheta}(t) &= -B_y^\top \vartheta(t) + P \hat{x}(t) + R_y \hat{y}(t) \\
\vartheta(T) &= -R_0 y(T) \\
0 &= B_x^\top \vartheta(t) - R_x \hat{v}(t) - D_v^\top \lambda(t) \\
0 &\leq \lambda(t) - (D_u \hat{u}(t) + D_v \hat{v}(t) - \delta(t))^\top \lambda(t) = 0.
\end{align*}
\]

Observe that the above optimality conditions (27) are an equivalent representation of the actual KKT-conditions of problem (26), called Pontryagin Maximum Principle. However, using these conditions to replace (25) by a single-level problem leads to a surrogate problem, which is not of the form (13) since we already exploited a reformulation of the lower level KKT-conditions. Hence, by applying the theory of this article it is possible to discuss two different approaches on how to deal with (25): an indirect approach which transforms (25) into an MPCC by means of Remark 5.1, and a direct approach making use of the results presented in Section 4 and Example 4.2, especially. Both approaches turn out to lead to the same necessary optimality conditions and constraint qualifications for (25). Hence, we just present the indirect approach here and leave it to the interested reader to verify the validity of these results evaluating the optimality system (16) for (25).

Due to Remark 5.1, we replace (25) by the following single-level optimal control problem with complementarity constraints:
\[
F_0(x(T), y(T)) + \int_0^T F_1(t, x(t), y(t), u(t), v(t)) \, dt \rightarrow \min_{x,y,u,v,\vartheta,\lambda,\xi} \quad \text{s.t.} \quad \begin{align*}
\dot{x}(t) &= C_x x(t) + D_x u(t) + \vartheta(t) \\
\dot{y}(t) &= C_y y(t) + D_y v(t) \\
\dot{\vartheta}(t) &= A_y \vartheta(t) + B_y \vartheta(t) \\
x(0) - x_0 &= 0 \\
y(0) - y_0 &= 0 \\
\vartheta(0) - \xi &= 0 \\
\dot{\vartheta}(T) + R_0 y(T) &= 0 \\
B_x^\top \vartheta(t) - R_x v(t) - D_v^\top \lambda(t) &= 0 \\
D_u u(t) + D_v v(t) - \delta(t) &\leq 0 \\
-\lambda(t) &\leq 0 \\
\int_0^T (D_u u(t) + D_v v(t) - \delta(t))^\top \lambda(t) \, dt &= 0.
\end{align*}
\]

Note that the complementarity condition from Remark 5.1 is replaced w.l.o.g. by an integral since the latter construction equals the dual pairing in the space \( L_2^n(0,T) \). Furthermore, we added the dummy variable \( \xi \in \mathbb{R}^m \) for the purpose of constructing a suitable constraint qualification for (28) later on.

Now we introduce the reflexive Banach spaces
\[
X := W_{1,2}^n(0,T) \times W_{1,2}^m(0,T) \times W_{1,2}^m(0,T) \times L_2^k(0,T) \times L_2^l(0,T) \times L_2(0,T) \times \mathbb{R}^m \\
W := W_{1,2}^n(0,T) \times W_{1,2}^m(0,T) \times W_{1,2}^m(0,T) \times \mathbb{R}^m \times L_2(0,T) \\
Z := L_2(0,T)
\]
and the cones $K := \{ v(t) \}$ as well as
\[
C := \{ w \in L^2(0, T) \mid w(t) \leq 0 \text{ for a.e. } t \in (0, T) \}. 
\]

We identify the Hilbert space $Z$ with its dual space, and this implies $C^\circ \cong -C$. Let us define $a : X \to \mathbb{R}$, $b : X \to W$, $c : X \to Z$, and $d : X \to Z^\ast$ for any $p := (x, y, \vartheta, u, v, \lambda, \xi) \in X$ by
\[
a(p) := F_0(x(T), y(T)) + \int_0^T F_1(t, x(t), y(t), u(t), v(t)) \, dt,
\]
\[
b(p) := \left( x(\cdot) - x_0 - \int_0^\cdot \left[ C_x x(\tau) + C_u u(\tau) \right] \, d\tau, \right.
\]
\[
y(\cdot) - y_0 - \int_0^\cdot \left[ A_x x(\tau) + B_y y(\tau) + A_u u(\tau) + B_v v(\tau) \right] \, d\tau,
\]
\[
\partial(\cdot) - \xi - \int_0^\cdot \left[ \mathcal{P} x(\tau) + \mathcal{R}_y y(\tau) - \mathcal{B}_v v(\tau) \right] \, d\tau,
\]
\[
d(\cdot) := \mathcal{D}_u u(\cdot) + \mathcal{D}_v v(\cdot) - \delta(\cdot).
\]

The latter is an affine bounded operator and the same holds true for $f'(p)$, $c'(p)$, and $d'(p)$. We identify the spaces $W^\sigma_1(0, T)$ and $\mathbb{R}^\sigma \times L^2_2(0, T)$ for any $\sigma \in \mathbb{N}^+$ with each other since it is possible to define a bijection $\Phi : W^\sigma_1(0, T) \to \mathbb{R}^\sigma \times L^2_2(0, T)$ as stated below:
\[
\forall g \in W^\sigma_1(0, T) : \quad \Phi(g) := (g(0), \bar{g}).
\]

The space $W^\sigma_1(0, T)$ is isomorphic to its dual space, and the corresponding dual pairing (or, equivalently, the corresponding inner product of this Hilbert space) can be defined as follows:
\[
\forall g = (g^\ast, g^\dagger), h = (h^\ast, h^\dagger) \in W^\sigma_1(0, T) : \quad \langle g, h \rangle = g^\ast h^\ast + \int_0^T g^\dagger(t)^\top h^\dagger(t) \, dt.
\]

Forthwith, we keep the above notation, i.e. any element $h \in W^\sigma_1(0, T)$ will be identified with the pair $(h^\ast, h^\dagger) := (h(0), \bar{h}) = \Phi(h) \in \mathbb{R}^\sigma \times L^2_2(0, T)$. Consequently, we have
\[
X \cong \mathbb{R}^m \times L^2_2(0, T) \times \mathbb{R}^m \times L^2_2(0, T) \times \mathbb{R}^m \times L^2_2(0, T) \times \mathbb{R}^m \times L^2_2(0, T) \times \mathbb{R}^m \times L^2_2(0, T) \times \mathbb{R}^m \times L^2_2(0, T)
\]
\[
W \cong \mathbb{R}^m \times L^2_2(0, T) \times \mathbb{R}^m \times L^2_2(0, T) \times \mathbb{R}^m \times L^2_2(0, T) \times \mathbb{R}^m \times L^2_2(0, T) \times \mathbb{R}^m \times L^2_2(0, T)
\]
and \( X \cong X^* \) as well as \( W \cong W^* \). Firstly, using integration by parts yields
\[
a'(p) = \left( \nabla_x F_0(x(T), y(T))^\top + \int_0^T \nabla_x F_1(\tau, x(\tau), y(\tau), u(\tau), v(\tau))^\top \, d\tau, \right. \\
\left. \nabla_y F_0(x(T), y(T))^\top + \int_0^T \nabla_y F_1(\tau, x(\tau), y(\tau), u(\tau), v(\tau))^\top \, d\tau, \right. \\
\left. \nabla_y F_0(x(T), y(T))^\top + \int_0^T \nabla_y F_1(\tau, x(\tau), y(\tau), u(\tau), v(\tau))^\top \, d\tau, \right. \\
\left. 0, \phi_{L^2(0,T)}, \nabla_y F_1(\cdot, x(\cdot), y(\cdot), u(\cdot), v(\cdot))^\top, \nabla_y F_1(\cdot, x(\cdot), y(\cdot), u(\cdot), v(\cdot))^\top, \right. \\
\left. \phi_{L^2(0,T)}, 0 \right)
\]
(cf. [8] for similar calculations). Now take \( \phi := (h,k,p,q,a) \in W^* \) and \( \phi \in X \) from above. Then from the definition of the adjoint operator we clearly have \( \langle \phi, (b'(p))^*[\phi] \rangle = \langle b'(p)[\phi], \phi \rangle \). In order to find an explicit representation of \( b'(p)^* \), we evaluate the right-hand side of this equation and apply integration by parts. We present this procedure by means of an easier operator in Appendix 6. Time-consuming calculations lead to the representation
\[
b'(p)^*[\phi] = \left( h^* - \int_0^T \left[ C^*_x h^f(\tau) + A^*_x k^f(\tau) + P^T p^f(\tau) \right] \, d\tau, \right. \\
\left. h^f(\cdot) - \int_0^T \left[ C^*_x h^f(\tau) + A^*_x k^f(\tau) + P^T p^f(\tau) \right] \, d\tau, \right. \\
\left. k^* + R_0 q - \int_0^T \left[ B^*_y k^f(\tau) + R_y p^f(\tau) \right] \, d\tau, \right. \\
\left. k^f(\cdot) + R_0 q - \int_0^T \left[ B^*_y k^f(\tau) + R_y p^f(\tau) \right] \, d\tau, \right. \\
\left. p^* + q + \int_0^T \left[ B_y p^f(\tau) + B_s s(\tau) \right] \, d\tau, \right. \\
\left. p^f(\cdot) + q + \int_0^T \left[ B_y p^f(\tau) + B_s s(\tau) \right] \, d\tau, \right. \\
\left. - C_q h^f(\cdot) - A^*_p k^f(\cdot) - B^*_p k^f(\cdot) - R_y s(\cdot), \right. \\
\left. - D_s s(\cdot), -p^* \right).
\]
Next, choose an arbitrary element \( \nu \in Z^* \). Then it is not difficult to obtain that
\[
c'(p)^*[\nu] = \left( 0, \phi_{L^2(0,T)}, 0, \phi_{L^2(0,T)}, 0, \phi_{L^2(0,T)}, D^*_u \nu(\cdot), D^*_u \nu(\cdot), \phi_{L^2(0,T)}, 0 \right)
\]
holds true. Similar as above we derive
\[
d'(p)^*[\mu] = \left( 0, \phi_{L^2(0,T)}, 0, \phi_{L^2(0,T)}, 0, \phi_{L^2(0,T)}, \phi_{L^2(0,T)}, \phi_{L^2(0,T)}, \mu(\cdot), 0 \right)
\]
for any \( \mu \in Z \).
Assume that $g := (h, k, p, q, s) \in W^*, \nu \in Z^*, \text{ and } \mu \in Z$ solve the system (8). Then, due to the above calculations, the following conditions are satisfied (for almost every $t \in (0, T)$):

$$
0 = \nabla_x F_0(x(T), y(T))^\top + h^x
+ \int_0^T \left[ \nabla_x F_1(t, x(t), y(t), u(t), v(t))^\top - C^x_2 h^x(t) - A^x_2 k^x(t) - P^T p^f(t) \right] \, dt,
\tag{30a}
$$

$$
0 = \nabla_y F_0(x(T), y(T))^\top + h^y
+ \int_0^T \left[ \nabla_y F_1(t, x(t), y(t), u(t), v(t))^\top - C^y_2 h^y(t) - A^y_2 k^y(t) - P^T p^f(t) \right] \, dt,
\tag{30b}
$$

$$
0 = \nabla_y F_0(x(T), y(T))^\top + k^y(t) + R_0 q
+ \int_0^T \left[ \nabla_y F_1(t, x(t), y(t), u(t), v(t))^\top - B^y_0 k^y(t) - R_y p^f(t) \right] \, dt,
\tag{30c}
$$

$$
0 = p^x + q + \int_0^T \left[ B_y p^x(t) + B_z s(t) \right] \, dt,
\tag{30d}
$$

$$
0 = p^y(t) + q + \int_0^T \left[ B_y p^y(t) + B_z s(t) \right] \, dt,
\tag{30e}
$$

$$
0 = \nabla_y F_1(t, x(t), y(t), u(t), v(t))^\top - C^y_u h^y(t) - A^y_u k^y(t) + D^y_u \nu(t),
\tag{30f}
$$

$$
0 = \nabla_y F_1(t, x(t), y(t), u(t), v(t))^\top - B^y_0 k^y(t) + D^y_0 \nu(t) - R_y \mu(t),
\tag{30g}
$$

$$
0 = -D_z s(t) + \mu(t),
\tag{30h}
$$

$$
0 = -p^y.
\tag{30i}
$$

From (30b), (30d), and (30f) we see that the functions $h^y$, $k^y$, and $p^y$ are elements of $W^y_{1,2}(0, T)$ (for appropriate $\sigma \in N^+$) again. Further, the equations (30a)–(30f) and (30j) yield that these functions are solutions of the boundary value problem (the differential equation shall hold for almost every $t \in (0, T)$)

$$
\begin{align*}
\dot{h}^x(t) &= \nabla_x F_1(t, x(t), y(t), u(t), v(t))^\top - C^x_2 h^x(t) - A^x_2 k^x(t) - P^T p^f(t),
\dot{k}^y(t) &= \nabla_y F_1(t, x(t), y(t), u(t), v(t))^\top - B^y_0 k^y(t) - R_y p^f(t),
p^x(0) = 0,
p^y(T) = -q,
\end{align*}
\tag{31}
$$

Moreover, (30g) and (30h) provide a linearized maximum condition (cf. (34)). Together with (31) and the equation $D_z s(\cdot) = \mu(\cdot)$ this system is equivalent to (30) and, hence, to

$$
0 \chi^* = a^*(p) + b^*(p)^* [\sigma] + c^*(p)^* [\nu] + d^*(p)^* [\mu].
$$

It remains to evaluate the sign conditions and complementarity conditions on the multipliers $g$, $\nu$, and $\mu$ from Definition 3.1. Let us introduce three sets $I^{+0}(i, p)$, $I^{00}(i, p)$, and $I^{0+}(i, p)$ by means of

$$
I^{+0}(i, p) := \{ t \in (0, T) \mid c_i(p)(t) < 0 \land d_i(p)(t) = 0 \}
$$

$$
I^{00}(i, p) := \{ t \in (0, T) \mid c_i(p)(t) = 0 \land d_i(p)(t) = 0 \}
$$

$$
I^{0+}(i, p) := \{ t \in (0, T) \mid c_i(p)(t) = 0 \land d_i(p)(t) > 0 \}
$$

for any $i \in \{1, \ldots, r\}$. Therein, $c_i(p), \ldots, c_r(p), d_i(p), \ldots, d_r(p) \in L_2(0, T)$ denote the component mappings of $c(p)$ and $d(p)$, respectively.
Lemma 5.1 Let \( p := (x, y, \vartheta, u, v, \lambda, \xi) \in X \) be a feasible point of (28). Then we have
\[
\begin{align*}
\cl(C^\circ - C \cap c(p)^\perp) \cap c(p)^\perp &= \{ \nu \in L^2_2(0,T) \mid \forall i \in \{1, \ldots, r\} : \nu_i(t) = 0 \ \text{for a.e.} \ t \in I^{10}(i,p) \} \\
\cl(C - C \cap d(p)^\perp) \cap d(p)^\perp &= \{ \mu \in L^2_2(0,T) \mid \forall i \in \{1, \ldots, r\} : \mu_i(t) = 0 \ \text{for a.e.} \ t \in I^{0+}(i,p) \}
\end{align*}
\]
as well as
\[
\begin{align*}
K_{C^\circ}(d(p), c(p)) &= \{ \nu \in L^2_2(0,T) \mid \forall i \in \{1, \ldots, r\} : \\
&\quad \nu_i(t) = 0 \ \text{for a.e.} \ t \in I^{10}(i,p) \} \\
K_{C}(c(p), d(p)) &= \{ \mu \in L^2_2(0,T) \mid \forall i \in \{1, \ldots, r\} : \\
&\quad \mu_i(t) = 0 \ \text{for a.e.} \ t \in I^{10+}(i,p) \}
\end{align*}
\]

Proof Since \( c(p)(t) \leq 0 \) holds for almost every \( t \in (0,T) \), we obtain
\[
C^\circ \cap c(p)^\perp = (C^\circ \cap c(p)) = -C \cap c(p)^\perp = \left\{ \nu \in -C \left| \sum_{i=1}^{r} \int_0^T \nu_i(t) c_i(p)(t) \ dt = 0 \right. \right\}
= \{ \nu \in -C \mid \forall i \in \{1, \ldots, r\} : \nu_i(t) = 0 \ \text{for a.e.} \ t \in I^{10}(i,p) \}.
\]
Hence, it is easy to see that
\[
C^\circ - C^\circ \cap c(p)^\perp = \{ \nu \in L^2_2(0,T) \mid \forall i \in \{1, \ldots, r\} : \nu_i(t) \geq 0 \ \text{for a.e.} \ t \in I^{10}(i,p) \}
\]
holds true and this set is closed. This yields the first assertion. It is well-known that the tangent cone \( T_{C^\circ}(d(p)) \) is given by means of
\[
T_{C^\circ}(d(p)) = \{ \nu \in L^2_2(0,T) \mid \forall i \in \{1, \ldots, r\} : \nu_i(t) \geq 0 \ \text{for a.e.} \ t \in I^{10}(i,p) \} \cup I^{00}(i,p),
\]
see [5, Example 2.64]. This implies the third assertion.
The second and fourth assertion follow analogously. \( \square \)

In the following lemma, we collect all the above results to introduce the notions of weak and strong stationarity in the sense of Definition 3.1 applied to (28).

Lemma 5.2 Let \( p := (x, y, \vartheta, u, v, \lambda, \xi) \in X \) be an arbitrary feasible point of (28).

1. The point \( p \) is weakly stationary for (28) if and only if there exist \( \phi_x \in W^{1,2}_T(0,T), \phi_y, \phi_\vartheta \in W^{1,2}_T(0,T), \) \( \nu \in L^2_2(0,T), \) and \( \beta \in L^2_2(0,T) \) as well as a vector \( \eta \in \mathbb{R}^m \) which satisfy the following conditions (for almost every \( t \in (0,T) \));
   (a) Adjoint condition
   \[
   \begin{align*}
   \dot{\phi}_x(t) &= -\nabla_x F_1(t, x(t), y(t), u(t), v(t))^\top - C^\top_x \phi_x(t) - A^\top_x \phi_y(t) - P^\top \phi_\vartheta(t) \\
   \dot{\phi}_y(t) &= -\nabla_y F_1(t, x(t), y(t), u(t), v(t))^\top - B^\top_y \phi_y(t) - R_y \phi_\vartheta(t) \\
   \dot{\phi}_\vartheta(t) &= B_y \phi_\vartheta(t) - B_u \beta(t),
   \end{align*}
   \]
   (b) Transversality condition
   \[
   \begin{align*}
   \phi_x(T) &= -\nabla_x F_0(x(T), y(T))^\top \\
   \phi_y(T) &= -\nabla_y F_0(x(T), y(T))^\top + R_0 \eta \\
   \phi_\vartheta(0) &= 0 \\
   \phi_\vartheta(T) &= \eta,
   \end{align*}
   \]
   (c) Linearized maximum condition
   \[
   \begin{align*}
   0 &= \nabla_u F_1(t, x(t), y(t), u(t), v(t))^\top - C_u \phi_x(t) - A_u^\top \phi_y(t) + D_u^\top \nu(t) \\
   0 &= \nabla_v F_1(t, x(t), y(t), u(t), v(t))^\top - B_v^\top \phi_y(t) + D_v^\top \nu(t) + R_v \phi_\vartheta(t),
   \end{align*}
   \]
(d) Extinction condition

\[ \forall i \in \{1, \ldots, r\}: \begin{cases} 
\nu_i(t) = 0 & \text{for a.e. } t \in I^+(i,p) \\
(D_v \beta(t))_i = 0 & \text{for a.e. } t \in I^+(i,p).
\end{cases} \tag{35} \]

2. The point \( p \) is strongly stationary for (28) if and only if there exist functions \( \phi_x \in W^m_{1,2}(0,T), \phi_y, \phi_\phi \in W^m_{1,2}(0,T), \nu \in L^2_2(0,T), \) and \( \beta \in L^2_2(0,T) \) as well as a vector \( \eta \in \mathbb{R}^n \) which satisfy (32), (33), (34), (35) and the following condition (for almost every \( t \in (0,T) \)):

(e) Nonnegativity condition

\[ \forall i \in \{1, \ldots, r\}: \begin{cases} 
\nu_i(t) \geq 0 & \text{for a.e. } t \in I^0(i,p). \\
(D_v \beta(t))_i \geq 0 & \text{for a.e. } t \in I^0(i,p).
\end{cases} \tag{36} \]

Proof Thanks to the calculations presented earlier this follows from Lemma 5.1, (31), and the corresponding linearized maximum conditions (30g) and (30h) by defining \( \phi_x := h^l, \phi_y := k^l, \phi_\phi := p^l, \beta := -s \) and \( \eta := -q \). We just replaced \( \mu(\cdot) \) by \(-D_v \beta(\cdot)\), while the negation of \( s \) provides consistent signs in the strong stationarity conditions. \( \square \)

Now we present a qualification condition, which ensures that a local optimal solution of (28) is strongly stationary. Therefore, we introduce the mapping \( H : X \rightarrow W \times Z \times Z^* \) by

\[ \forall p := (x, y, \vartheta, u, v, \lambda, \xi) \in X: \quad H(p) := (h(p), c(p), d(p)). \]

Since \( H \) is a bounded affine operator, it is continuously Fréchet differentiable at any point \( p \in X \), and the Fréchet derivative is independent of \( p \). Consequently, we can introduce \( \mathcal{H} \in L[X, W \times Z \times Z^*] \) to be the Fréchet derivative of \( H \) at an arbitrary point in \( X \). Observe that the following holds true for any \( \mathfrak{d} := (\mathfrak{d}_x, \mathfrak{d}_y, \mathfrak{d}_\vartheta, \mathfrak{d}_u, \mathfrak{d}_v, \mathfrak{d}_\lambda, \mathfrak{d}_\xi) \):

\[ \mathcal{H}[\mathfrak{d}] = \begin{pmatrix} 
\mathfrak{d}_x(\cdot) - \int_0^T \left[ C_x \mathfrak{d}_x(\tau) + C_u \mathfrak{d}_u(\tau) \right] d\tau, \\
\mathfrak{d}_y(\cdot) - \int_0^T \left[ A_x \mathfrak{d}_x(\tau) + B_y \mathfrak{d}_y(\tau) + A_u \mathfrak{d}_u(\tau) + B_v \mathfrak{d}_v(\tau) \right] d\tau, \\
\mathfrak{d}_\vartheta(\cdot) - \mathfrak{d}_\xi - \int_0^T \left[ P \mathfrak{d}_x(\tau) + R_u \mathfrak{d}_u(\tau) - B_v^\top \partial \phi(\tau) \right] d\tau, \\
\mathfrak{d}_\vartheta(T) + R_u \partial \vartheta(T), \\
B_v^\top \mathfrak{d}_\vartheta(\cdot) - R_u \mathfrak{d}_u(\cdot) - D^\top \mathfrak{d}_\lambda(\cdot), \\
D_u \mathfrak{d}_u(\cdot) + D_v \mathfrak{d}_v(\cdot), \\
\mathfrak{d}_\lambda(\cdot)
\end{pmatrix}. \]

Due to Theorem 3.1, the surjectivity of this operator implies that any local optimal solution of (28) is a strongly stationary point of that problem. Introducing a Banach space \( X' \) by

\[ X' := W^m_{1,2}(0,T) \times W^m_{1,2}(0,T) \times W^m_{1,2}(0,T) \times L^2_2(0,T) \times L^2_2(0,T) \times \mathbb{R}^n, \]
the operator $\mathcal{H}$ is surjective if and only if the operator $\tilde{\mathcal{H}} \in L[X', W \times Z]$ which is defined for arbitrary points $\delta := (\vartheta_x, \vartheta_y, \vartheta_\vartheta, \vartheta_v, \vartheta_\xi)$ by

$$\tilde{\mathcal{H}}[\delta] = \left( \vartheta_x(\cdot) - \int_0^\cdot \left[ C_x \vartheta_x(\tau) + C_u \vartheta_u(\tau) \right] \, d\tau, \right.$$  
$$\vartheta_y(\cdot) - \int_0^\cdot \left[ A_x \vartheta_x(\tau) + B_y \vartheta_y(\tau) + A_u \vartheta_u(\tau) + B_v \vartheta_v(\tau) \right] \, d\tau, \)

$$\vartheta_\vartheta(\cdot) - \vartheta_\xi = \int_0^\cdot \left[ P \vartheta_x(\tau) + R_y \vartheta_y(\tau) + B_x \vartheta_\vartheta(\tau) \right] \, d\tau,$$

$$\vartheta_v(\cdot) = \vartheta_\xi(\cdot) - R_v \vartheta_v(\cdot), \quad \vartheta_v(\cdot) = (\vartheta_v(0) + D_u \vartheta_u(\cdot)) = (\vartheta_v(0) + D_v \vartheta_v(\cdot)),$$

is surjective (cf. proof of Theorem 4.1 and Remark 4.2, where we exploited a similar reduction procedure).

Let us introduce $\hat{A} \in \mathbb{R}^{(n+2m)\times(n+2m)}$, $\hat{B} \in \mathbb{R}^{(n+2m)\times(k+l)}$, $\hat{C} \in \mathbb{R}^{(l+r)\times(n+2m)}$, $\hat{D} \in \mathbb{R}^{(l+r)\times(k+l)}$, $\tilde{H} \in \mathbb{R}^{m\times(n+2m)}$, and $K \in \mathbb{R}^{(n+2m)\times m}$ by means of

$$\hat{A} := \begin{pmatrix} C_x & O & O \\ A_x & B_y & O \\ P & R_y & -B_y^\top \end{pmatrix} \quad \hat{B} := \begin{pmatrix} C_u & O \\ A_u & B_v \\ O & O \end{pmatrix} \quad \hat{C} := \begin{pmatrix} O & O & B_x^\top \\ O & O & O \end{pmatrix} \quad \hat{D} := \begin{pmatrix} O & -R_v \\ D_u & D_v \end{pmatrix} \quad (\hat{H} := (O R_0 I) \quad \hat{K} := \begin{pmatrix} O \\ O \end{pmatrix}).$$

Using these matrices, we have the representation

$$\tilde{\mathcal{H}}[\delta] = \left( \vartheta_x(\cdot) - \int_0^\cdot \left( \hat{A} \begin{pmatrix} \vartheta_x(\tau) \\ \vartheta_y(\tau) \\ \vartheta_\vartheta(\tau) \end{pmatrix} + \hat{B} \begin{pmatrix} \vartheta_u(\tau) \\ \vartheta_v(\tau) \end{pmatrix} \right) \, d\tau, \right.$$  
$$\tilde{\mathcal{H}}[\delta] = \hat{H} \begin{pmatrix} \vartheta_x(T) \\ \vartheta_y(T) \\ \vartheta_\vartheta(T) \end{pmatrix} + \hat{C} \begin{pmatrix} \vartheta_v(T) \\ \vartheta_\xi(T) \end{pmatrix} + \hat{D} \begin{pmatrix} \vartheta_u(T) \\ \vartheta_v(T) \end{pmatrix}.$$

Taking an arbitrary point $\tilde{\gamma} := (\vartheta_x, \vartheta_y, \vartheta_\vartheta, \vartheta_v, \vartheta_\xi) \in W \times Z$, we are going to formulate a criterion which ensures that the equation $\tilde{\mathcal{H}}[\delta] = \tilde{\gamma}$ possesses a solution.

We start with the last two components of the latter equation, which are given by

$$\begin{align*}
\hat{C} \begin{pmatrix} \vartheta_v(T) \\ \vartheta_\xi(T) \end{pmatrix} + \hat{D} \begin{pmatrix} \vartheta_u(T) \\ \vartheta_v(T) \end{pmatrix} &= \begin{pmatrix} \vartheta_w(T) \\ \vartheta_\xi(T) \end{pmatrix}.
\end{align*}$$

(38)

Assume that $\hat{D} \in \mathbb{R}^{(l+r)\times(k+l)}$ possesses full row rank $l + r$ and $r < k$ holds true. Then there exists a matrix $\hat{Y} \in \mathbb{R}^{(l+r)\times(k-r)}$ possessing full column rank $k - r$ which satisfies $\hat{D} \hat{Y} = 0$ (the columns of $\hat{Y}$ form a basis of the nullspace of $\hat{D}$). Introducing the pseudo inverse matrix $\hat{D}^\dagger := \hat{D}^\dagger (\hat{D} \hat{D}^\dagger)^{-1}$ of $\hat{D}$, we easily see that (38) can be transformed into

$$\begin{align*}
\begin{pmatrix} \vartheta_u(T) \\ \vartheta_v(T) \end{pmatrix} &= \hat{Y} \omega(\cdot) + \hat{D}^\dagger \begin{pmatrix} \vartheta_w(T) \\ \vartheta_\xi(T) \end{pmatrix} - \hat{C} \begin{pmatrix} \vartheta_v(T) \\ \vartheta_\xi(T) \end{pmatrix}.
\end{align*}$$

for any function $\omega \in L_2^{l-r}(0, T)$. Hence, we can reduce solving $\tilde{\mathcal{H}}[\delta] = \tilde{\gamma}$ to the problem of finding $(\vartheta_x, \vartheta_y, \vartheta_\vartheta, \vartheta_v, \vartheta_\xi, \omega) \in W_1^{m}(0, T) \times W_1^{m}(0, T) \times W_1^{m}(0, T) \times W_2^{m}(0, T) \times L_2^{l-r}(0, T)$ (39)
which satisfies
\[
\begin{pmatrix}
\hat{\theta}_x(t) \\
\hat{\theta}_y(t) \\
\hat{\theta}_\alpha(t)
\end{pmatrix} = \begin{pmatrix}
\hat{\theta}_x(t) \\
\hat{\theta}_y(t) \\
\hat{\theta}_\alpha(t)
\end{pmatrix} + \hat{B}Y\omega(t)
\]
and the boundary conditions
\[
\begin{pmatrix}
\hat{\theta}_x(0) \\
\hat{\theta}_y(0) \\
\hat{\theta}_\alpha(0)
\end{pmatrix} = \begin{pmatrix}
\theta_x(0) \\
\theta_y(0) \\
\theta_\alpha(0)
\end{pmatrix}
\text{ and } \begin{pmatrix}
\hat{\theta}_x(T) \\
\hat{\theta}_y(T) \\
\hat{\theta}_\alpha(T)
\end{pmatrix} = g_s.
\]

Since the system is linear, we can eliminate the inhomogeneity from the differential equation and the initial condition. It remains to guarantee, that for every \(g_s \in \mathbb{R}^m\) we find a solution \((39)\) of
\[
\begin{pmatrix}
\hat{\theta}_x(t) \\
\hat{\theta}_y(t) \\
\hat{\theta}_\alpha(t)
\end{pmatrix} = \begin{pmatrix}
\hat{\theta}_x(t) \\
\hat{\theta}_y(t) \\
\hat{\theta}_\alpha(t)
\end{pmatrix} + \hat{B}Y\omega(t)
\]
\[
\begin{pmatrix}
\hat{\theta}_x(0) \\
\hat{\theta}_y(0) \\
\hat{\theta}_\alpha(0)
\end{pmatrix} = \begin{pmatrix}
\theta_x(0) \\
\theta_y(0) \\
\theta_\alpha(0)
\end{pmatrix}
\text{ and } \begin{pmatrix}
\hat{\theta}_x(T) \\
\hat{\theta}_y(T) \\
\hat{\theta}_\alpha(T)
\end{pmatrix} = g_s.
\]

By the full row rank \(m\) of \(\hat{H}\) and by fixing \(\hat{\theta}_\xi = 0\), a sufficient condition is the controllability (cf. [2]) for an introduction to the theory of dynamical systems and some theoretical analysis of their behaviour) of the following system \([17, \text{ proof of Theorem 5.19}]:\)
\[
\begin{pmatrix}
\hat{x}(t) \\
\hat{y}(t) \\
\hat{\theta}(t)
\end{pmatrix} = \begin{pmatrix}
\hat{A} - \hat{B}\hat{D}^+\hat{C} \\
\hat{A} - \hat{B}\hat{D}^+\hat{C} \\
\hat{A} - \hat{B}\hat{D}^+\hat{C}
\end{pmatrix} \begin{pmatrix}
\hat{\theta}_x(t) \\
\hat{\theta}_y(t) \\
\hat{\theta}_\alpha(t)
\end{pmatrix} + \hat{B}Y\omega(t).
\]

By the famous Kalman-Theorem [2, Theorem 4.1] the system \((41)\) is controllable if and only if its controllability matrix
\[
\begin{bmatrix}
\hat{B}Y \quad (\hat{A} - \hat{B}\hat{D}^+\hat{C})\hat{B}Y \\
\ldots \\
(\hat{A} - \hat{B}\hat{D}^+\hat{C})^{n+2m-1}\hat{B}Y
\end{bmatrix} \in \mathbb{R}^{(n+2m) \times (n+2m) (k-r)}
\]}

possesses full row rank \(n + 2m\).

**Lemma 5.3** Let \((x, u, v, v) \in W_{1,2}^m(0, T) \times L^2_0(0, T) \times W_{1,2}^m(0, T) \times L^2_0(0, T)\) be a local optimal solution of the bilevel programming problem \((25)\). Assume that the matrix \(\hat{D}\) possesses full row rank \(l + r\) and that one of the following conditions holds.

1. For every \(g_s \in \mathbb{R}^m\), the system \((40)\) has a solution \((39)\).

2. The controllability matrix from \((42)\) possesses full row rank \(n + 2m\).

Then there exist functions \(\phi_x \in W_{1,2}^m(0, T), \phi_y, \phi_\theta, \theta \in W_{1,2}^m(0, T), \nu, \lambda \in L^2_0(0, T)\), and \(\beta \in L^2_0(0, T)\) as well as a vector \(\eta \in \mathbb{R}^m\) which satisfy the conditions \((27), (32), (33), (34), (35),\) and \((36)\).

**Proof** Due to Remark 5.1, it is possible to find \(\theta \in W_{1,2}^m(0, T)\) and \(\lambda \in L^2_0(0, T)\) satisfying the conditions in \((27)\). One may check that \((x, y, \theta, u, \nu, \lambda, \vartheta(0)) \in X\) is a local optimal solution of \((28)\), which is in fact a direct consequence of Remark 5.1 again. Recalling the above considerations, the operator \(H\) is surjective, which means that \((x, y, \theta, u, \nu, \lambda, \vartheta(0))\) is a strongly stationary point of \((28)\) by Theorem 3.1. Consequently, by means of Lemma 5.2, we obtain the statement of this lemma. \(\Box\)

As mentioned at the beginning of this section, \((25)\) is a special case of the general bilevel programming problem we already considered in Example 4.2. Applying \((16)\) to \((25)\) yields the following result.

**Lemma 5.4** Let \((x, u, y, v) \in X \times Y\) be an arbitrary feasible point of the bilevel optimal control problem \((25)\). Then it is strongly stationary (in the sense of Definition 4.1) for \((25)\) if and only if there exist \(\phi_x \in W_{1,2}^m(0, T), \phi_y, \phi_\theta, \theta \in W_{1,2}^m(0, T), \nu, \lambda \in L^2_0(0, T),\) and \(\beta \in L^2_0(0, T)\) as well as a vector \(\eta \in \mathbb{R}^m\) which satisfy the conditions \((27), (32), (33), (34), (35),\) and \((36)\).
Hence, combining Lemmata 5.3 and 5.4 yields our final result.

**Theorem 5.1** Let \((\hat{x}, \hat{u}, \hat{y}, \hat{v}) \in W^1_{1,2}(0, T) \times L^2_2(0, T) \times W^m_{1,2}(0, T) \times L^2_2(0, T)\) be a local optimal solution of the bilevel programming problem (25). Assume that the matrix \(\hat{D}\) possesses full row rank \(l + r\) and that one of the following conditions holds.

1. For every \(g_n \in \mathbb{R}^m\), the system (40) has a solution (39).
2. The controllability matrix from (42) possesses full row rank \(n + 2m\).

Then \((\hat{x}, \hat{u}, \hat{y}, \hat{v})\) is strongly stationary (in the sense of Definition 4.1) for (25).

### 6 Appendix

For natural numbers \(k_1, k_2, k_3 \in \mathbb{N}^+\) as well as matrices \(M_1 \in \mathbb{R}^{k_1 \times k_1}, M_2 \in \mathbb{R}^{k_1 \times k_2}, M_3 \in \mathbb{R}^{k_3 \times k_2}\), we consider the bounded linear operator

\[
\mathcal{D} \in L[W^k_{1,2}(0, T) \times L^2_2(0, T), W^{k_2}_{1,2}(0, T) \times L^2_2(0, T)]
\]

given by

\[
\mathcal{D}[z, w] := \left( z(\cdot) - \int_0^T [M_1 z(\tau) + M_2 w(\tau)] \, d\tau, M_3 w(\cdot) \right)
\]

for any \(z \in W^k_{1,2}(0, T)\) and \(w \in L^2_2(0, T)\).

We apply the definitions of the inner product in \(W^k_{1,2}(0, T)\) and \(L^2_2(0, T)\) as well as integration by parts in order to derive the following equality for arbitrary functions \(z \in W^k_{1,2}(0, T)\) and \(v \in L^2_2(0, T)\):

\[
((z, w), \mathcal{D}^*[z, v]) = \langle \mathcal{D}[z, w], (z, v) \rangle
\]

\[
= z^* z^T + \int_0^T \left( z^f(\tau) - M_1 z^T (z^* + \int_0^T z^f(s) \, ds) - M_2 w(\tau) \right)^T z^f(\tau) \, d\tau
\]

\[
+ \int_0^T (M_3 w(\tau))^T v(\tau) \, d\tau
\]

\[
= z^* z^T \left( z^* - \int_0^T M_1^T z^f(\tau) \, d\tau \right) + \int_0^T w(\tau)^T \left( M_3^T v(\tau) - M_2^T z^f(\tau) \right) \, d\tau
\]

\[
+ \int_0^T z^f(\tau)^T z^f(\tau) - \left( \int_0^T z^f(s) \, ds \right)^T M_1^T z^f(\tau) \, d\tau
\]

\[
= z^* z^T \left( z^* - \int_0^T M_1^T z^f(\tau) \, d\tau \right) + \int_0^T w(\tau)^T \left( M_3^T v(\tau) - M_2^T z^f(\tau) \right) \, d\tau
\]

\[
+ \int_0^T z^f(\tau)^T z^f(\tau) \, d\tau - \left( \int_0^T z^f(s) \, ds \right)^T \left( \int_0^T M_1^T z^f(s) \, ds \right) \bigg|_0^T
\]

\[
+ \int_0^T z^f(\tau)^T \left( \int_0^\tau M_1^T z^f(s) \, ds \right) \, d\tau
\]

\[
= z^* z^T \left( z^* - \int_0^T M_1^T z^f(\tau) \, d\tau \right) + \int_0^T w(\tau)^T \left( M_3^T v(\tau) - M_2^T z^f(\tau) \right) \, d\tau
\]

\[
+ \int_0^T z^f(\tau)^T z^f(\tau) \, d\tau - \int_0^T z^f(\tau)^T \left( \int_0^\tau M_1^T z^f(s) \, ds \right) \, d\tau
\]

\[
= z^* z^T \left( z^* - \int_0^T M_1^T z^f(\tau) \, d\tau \right) + \int_0^T z^f(\tau)^T \left( z^f(\tau) - \int_0^T M_1^T z^f(s) \, ds \right) \, d\tau
\]

\[
+ \int_0^T w(\tau)^T \left( M_3^T v(\tau) - M_2^T z^f(\tau) \right) \, d\tau.
\]
Therein, we made use of the fact that $W_{1,2}^k(0,T)$ can be decomposed to the space $\mathbb{R}^{k_1} \times L^2_{\mathcal{D}}(0,T)$ as described in Section 5. Now it is easy to see from the definition of the dual pairing that the adjoint operator $D^*$ of $\mathcal{D}$ can be represented by

$$D^*[z,v] = \left( z^* - \int_0^T M_1^T z^f(\tau) d\tau, z^f(\cdot) - \int_0^T M_1^T z^f(\tau) d\tau, M_3^* v(\cdot) - M_2^* z^f(\cdot) \right)$$

for any $z \in W_{1,2}^k(0,T)$ and $v \in L^2_{\mathcal{D}}(0,T)$.

References