Bilevel optimal control with final-state-dependent finite-dimensional lower level
Francisco Benita, Patrick Mehlitz

Bilevel optimal control with final-state-dependent finite-dimensional lower level
ISSN 1433 – 9307

Herausgeber: Dekan der Fakultät für Mathematik und Informatik

Herstellung: Medienzentrum der TU Bergakademie Freiberg
Bilevel optimal control with final-state-dependent finite-dimensional lower level

Francisco Benita¹, Patrick Mehlitz²,*

Abstract
In this paper we discuss special bilevel optimal control problems where the upper level problem is an optimal control problem of ODEs with control and terminal constraints and the lower level problem is a finite-dimensional parametric optimization problem where the parameter is the final state of the state variable of the upper level. We tackle this problem using tools of nonsmooth analysis, optimization in Banach spaces and bilevel programming to derive a necessary optimality condition of linearized Pontryagin-type.

Keywords: Optimal control, Bilevel programming, Pontryagin Maximum Principle, Optimization in Banach spaces, Nonsmooth optimization, Calmness

1. Introduction and notations
In this paper we consider special optimistic bilevel optimal control problems with an optimal control problem in the upper level (problem of the leader) and a parametric but finite-dimensional optimization problem in the lower level (problem of the follower) whose parameter is nothing else but the final state of the leader’s state variable. To be more exact we have

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

\[ \dot{x}(t) = F(t, x(t), u(t)) \quad \text{a.e. } t \in [0, T] \]

\[ x(0) = x_0 \]

\[ G(t, u(t)) \leq a_{R^p} \quad \text{a.e. } t \in [0, T] \]

\[ H(x(T)) \leq a_{R^s} \]

\[ y \in \Psi(x(T)) \]

where \( \Psi: \mathbb{R}^n \rightarrow \mathcal{P}(\mathbb{R}^k) \) is the solution set mapping of:

\[ f(z, y) \rightarrow \min_y \]

\[ g(z, y) \leq a_{R^p} \]

\[ h(z, y) = a_{R^s}. \]

Henceforth, we suppose that the following assumptions hold:

- \( F_0: \mathbb{R}^n \times \mathbb{R}^k \rightarrow \mathbb{R}, F_1: \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^k \rightarrow \mathbb{R}, F: \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n, G: \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}', H: \mathbb{R}^n \rightarrow \mathbb{R}^s, f: \mathbb{R}^n \times \mathbb{R}^k \rightarrow \mathbb{R}, g: \mathbb{R}^n \times \mathbb{R}^k \rightarrow \mathbb{R}^p \) and \( h: \mathbb{R}^n \times \mathbb{R}^k \rightarrow \mathbb{R}^q \) are continuously differentiable functions.

*Corresponding author
¹ITESM (Tec de Monterrey), Campus Monterrey, Mexico, francisco_benita@hotmail.com
²TU Bergakademie Freiberg, Germany, mehlitz@math.tu-freiberg.de
the final time $T > 0$ and the initial state $x_0 \in \mathbb{R}^m$ at $t = 0$ are fixed,

- the state function $x$ and the control function $u$ are chosen from the spaces $W_{1,\infty}^m[0, T]$ and $L_{\infty}^m[0, T]$, respectively.

We are going to derive necessary optimality conditions of Pontryagin-type for the problem (1), (2) using ideas from the theory of bilevel programming, nonsmooth analysis, optimization in Banach spaces, ordinary differential equations and optimal control. Therefore, we proceed as follows: After we have introduced some basic notations in the first chapter we derive a necessary optimality condition of Karush-Kuhn-Tucker-type (in the following we use the abbreviation KKT) for general optimization problems in Banach spaces with a locally Lipschitz continuous objective functional and continuously Fréchet differentiable constraints. We obtain the desired result dealing with the well-known constraint qualification of Kurczyusz, Robinson and Zowe introduced in [14, 18] and using the Clarke generalized derivative. In Section 3 we give a short introduction to bilevel programming. Afterwards, considering the optimal value reformulation of the original problem (1), (2), we introduce the well-known concept of partial calmness which was formulated by Ye and Zhu in [21] first. This property of the problem allows us to shift the crucial constraint containing the optimal value function of the lower level problem to the upper level objective function; in fact, this new problem is nothing else but a partially penalized optimal value reformulation. Furthermore, we give some conditions under which the partial calmness property holds. Especially, we take a closer look at the so-called value-function-constraint-qualification of Henrion and Surowiec (cf. [12]). Finally, in Section 4 we derive a necessary optimality condition of linearized Pontryagin-type for problem (1), (2) by constructing a necessary optimality condition for the partially penalized optimal value reformulation which is a special optimal control problem with mixed and terminal constraints again. In order to ensure the regularity of the latter problem we introduce and analyse a constraint qualification called upper level regularity. Moreover, we give examples to verify that the derived theory is applicable.

The idea of applying the theory of optimization problems in Banach spaces in order to derive necessary optimality conditions for optimal control problems had already been used by Gerdts (cf. [11]), Jahn (cf. [13]) and many other authors. However, we are going to tackle the above bilevel programming problem with these tools and to our knowledge there is no literature available where similar research had been done.

First of all we are going to introduce several concepts of optimization and variational analysis as well as some notations we want to use throughout this paper. Forthwith, let $\mathbb{N}$, $\mathbb{R}$, $\mathbb{R}^n$, $\mathbb{R}^{m\times n}$, $\mathbb{R}^{n,*}$ and $\mathbb{R}^+$ denote the natural numbers with zero, the real numbers, the extended real numbers (i.e. $\overline{\mathbb{R}} = \mathbb{R} \cup [-\infty, +\infty]$), the space of all vectors with $n$ components, the set of all real matrices with $m$ rows and $n$ columns, the cone of all vectors in $\mathbb{R}^n$ with non-negative components and the set of all vectors in $\mathbb{R}^n$ with positive components. Furthermore, we postulate that the relations $<, \leq$ and $=$ should be interpreted componentwise if the compared objects are vectors of the same dimension. If $M \in \mathbb{R}^{m\times n}$ is an arbitrarily chosen matrix, then $M^T \in \mathbb{R}^{n\times m}$ is used to express its transpose. Forthwith, we use $I_n \in \mathbb{R}^{n\times n}$ and $O^{m\times n} \in \mathbb{R}^{m\times n}$ to express the identity matrix and the zero matrix of appropriate dimensions. Let $X$ be a real Banach space with its corresponding norm $\| \cdot \|_X$ and zero vector $0_X$. For a nonempty set $A \subseteq X$ the sets $\text{int}(A)$, $\text{cl}(A)$ and $\text{cone}(A)$ denote the interior, the closure and the conic hull of $A$, respectively. Furthermore, $X'$ shall represent the dual space of $X$ and $\langle \cdot, \cdot \rangle : X \times X' \rightarrow \mathbb{R}$ is used to express the corresponding dual pairing. Additionally, we use $A^D = \{ p \in X' | x \in A : \langle x, p \rangle \geq 0 \}$ for the dual cone of $A$ and if $A$ is convex $N_A(\bar{x}) = -A - \{ \bar{x} \}^D$ for the normal cone to $A$ at $\bar{x} \in A$, respectively. Let $\mathbb{B}_X$ and $U_X$ be the closed and open unit ball of $X$, respectively. For any $\varepsilon > 0$ and $\bar{x} \in X$ we define $U_\varepsilon^\bar{x}(\bar{x}) := \{ \bar{x} \} + \varepsilon \cdot U_X$. Let $Y$ and $Z$ be two real Banach spaces and $F : X \rightarrow Y$ and $G : Y \rightarrow Z$ bounded linear operators. We denote the kernel of $F$ by $\ker(F) = \{ x \in X | F(x) = 0 \}$ and the composition of $F$ and $G$ by $(F \circ G)(x) = G(F(x))$ for any $x \in X$. Henceforth, we define the norm in the product space $X \times Y$ by $\|(x, y)\|_{X \times Y} = \max\{\|x\|_X, \|y\|_Y\}$ for any pair $(x, y) \in X \times Y$. If not mentioned otherwise, we equip the space $\mathbb{R}^n$ with the maximum-norm $\| \cdot \|_\infty$. Let $[0, T] \subseteq \mathbb{R}$ be a nonempty interval. Then we use $L_{\infty}^m[0, T]$ to express the space of all functions $x : [0, T] \rightarrow \mathbb{R}^m$ with essentially bounded component functions equipped with the maximum norm w.r.t. the $L_{\infty}^m[0, T]$-norms of all component functions. Similarly, we define the Banach space $L_{\infty}^n[0, T]$ whose elements are functions possessing $m$ components which belong to $L_{1}[0, T]$. The Banach space $W_{1,\infty}^m[0, T]$ contains all functions $x : [0, T] \rightarrow \mathbb{R}^m$ possessing an essentially bounded weak derivative $\dot{x}$ equipped with the following norm:

$$\forall x \in W_{1,\infty}^m[0, T]: \quad \|x\|_{W_{1,\infty}^m[0, T]} := \max\{\|x\|_{L_{\infty}^m[0, T]}, \|\dot{x}\|_{L_{\infty}^m[0, T]}\}.$$ 

Moreover, we use $W_{1,\infty}^m[0, T]$ to express the space of all functions $x : [0, T] \rightarrow \mathbb{R}^m$ whose component functions come from $W_{1,\infty}^m[0, T]$. We equip this space again with the maximum norm w.r.t. the $W_{1,\infty}^m[0, T]$-norms of all component
functions. Henceforth, we use the Banach spaces $E_1$, $E_2$ and $E_3$ defined as follows:

$$E_1 := W_{1,\infty}^0[0, T] \times L^\infty_{\text{loc}}[0, T]$$

$$E_2 := E_1 \times \mathbb{R}^k$$

$$E_3 := E_2 \times \mathbb{R}.$$ 

Let $h_1: X \rightarrow Y$, $h_2: \mathbb{R}^n \times \mathbb{R}^k \rightarrow \mathbb{R}^m$ and $\psi: X \rightarrow \mathbb{R}$ be Fréchet differentiable, differentiable and locally Lipschitz continuous, respectively. Then for any points $\hat{x} \in X$ and $(\hat{z}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^m$ let $h'_1(\hat{x})$, $\nabla h_2(\hat{z}, \hat{y})$ and $\nabla_2 h_2(\hat{z}, \hat{y})$ denote the Fréchet derivative of $h_1$ at $\hat{x}$, the Jacobian of $h_2$ at $(\hat{z}, \hat{y})$ and the partial Jacobian of $h_2$ at $(\hat{z}, \hat{y})$ w.r.t. $z$. Furthermore, for any $\hat{x}, d \in X$ we define the upper Clarke generalized derivative of $\psi$ at $\hat{x}$ in direction $d$ as follows:

$$\psi'_+(\hat{x}; d) := \limsup_{x \rightarrow \hat{x}, t \rightarrow 0} \frac{\psi(x + td) - \psi(x)}{t}.$$ 

Additionally, the set $\\partial \psi(\hat{x}) = \{ p \in X^* | \forall d \in X; \langle d, p \rangle \leq \psi'_+(\hat{x}; d) \}$ is called Clarke subdifferential of $\psi$ at $\hat{x}$. Since $\psi$ is locally Lipschitz continuous, this set is nonempty, bounded, closed and convex (cf. [6]). If a function $\theta: \mathbb{R} \rightarrow \mathbb{R}^n$ is (weakly) differentiable, we denote its (weak) derivative by $\dot{\theta}$.

Let $\Theta: \mathbb{R}^n \rightarrow \mathcal{P}(\mathbb{R}^k)$ be a set-valued mapping ($\mathcal{P}(\cdot)$ denotes the power set operator). Then we define its domain and graph by $\text{dom}(\Theta) = \{ z \in \mathbb{R}^n | \Theta(z) \neq \emptyset \}$ and $\text{graph}(\Theta) = \{(z, y) \in \mathbb{R}^n \times \mathbb{R}^k | y \in \Theta(z) \}$, respectively. We call $\Theta$ inner semicontinuous at some point $(\hat{z}, \hat{y}) \in \text{graph}(\Theta)$ if for any sequence $(z_l) \subseteq \mathbb{R}^n$ converging to $\hat{z}$ there exist an index $l_0 \in \mathbb{N}$ and a sequence $(y_l) \subseteq \mathbb{R}^k$ converging to $\hat{y}$ such that $y_l \in \Theta(z_l)$ holds for any $l \geq l_0$ (cf. [8]). Furthermore, $\Theta$ is called calm at $(\hat{z}, \hat{y})$ if there exist neighbourhoods $V$ of $\hat{z}$ and $W$ of $\hat{y}$ as well as some constant $L > 0$ such that

$$\forall z \in V; \quad \Theta(z) \cap W \subseteq \Theta(\hat{z}) + L \cdot \| z - \hat{z} \|_\infty \cdot \mathbb{B}_{\mathbb{R}^k}$$

holds true (cf. [12]). Observe that $\Theta$ is calm at $(\hat{z}, \hat{y})$ if it is locally Lipschitz-like (has the Aubin property) at this point (cf. [15]).

2. Generalized optimization problems with nonsmooth objective functional

In this section we are going to discuss the optimization problem

$$\psi(x) \rightarrow \min$$

$$\alpha(x) \in -K$$

$$\beta(x) = \sigma_Z$$

$$x \in \Omega$$

under the following assumptions:

- $X, Y$ and $Z$ are Banach spaces,
- $\psi: X \rightarrow \mathbb{R}$ is a locally Lipschitz continuous functional,
- $\alpha: X \rightarrow Y$ and $\beta: X \rightarrow Z$ are continuously Fréchet differentiable mappings,
- $K \subseteq Y$ is an ordering cone, i.e. a nonempty, closed, convex and pointed cone,
- $\Omega \subseteq X$ is a nonempty, closed and convex set.

Furthermore, we denote by $M = \{ x \in \Omega | \alpha(x) \in -K, \beta(x) = \sigma_Z \}$ the feasible set of problem (3). In order to derive necessary optimality conditions of KKT-type for this optimization problem one has to guarantee that some constraint qualification holds at the local optimal solutions of this problem. We make use of the well-known Kurcyusz-Robinson-Zowe-Constraint-Qualification (cf. [5, 9, 13, 14, 18]).
**Definition 2.1.** (KRZCQ)

Let \( \hat{x} \in M \) be a feasible point of problem (3). Then KRZCQ, the Kurcyusz-Robinson-Zowe-Constraint-Qualification, holds at \( \hat{x} \) provided we have:

\[
\partial Y_{\hat{x}} \in \text{int} \left( \left( [\alpha(\hat{x}), \partial_Z] + (\alpha'(\hat{x}), \beta'(\hat{x}))\{\Omega - \{\hat{x}\}] + K \times \{\partial_Z} \right) \right). \tag{4}
\]

Note that there exist several different possibilities to define KRZCQ in equivalent ways. Hence, the following lemma is included to give a short overview.

**Lemma 2.1.** [5, 9, 13, 18]

Let \( \hat{x} \in M \) be chosen arbitrarily. Then the following statements are equivalent:

(i) KRZCQ holds at \( \hat{x} \).

(ii) \( Y \times Z = (\alpha'(\hat{x}), \beta'(\hat{x}))[\text{cone}(\Omega - \{\hat{x}\})] + \text{cone}(K + [\alpha(\hat{x})]) \times \{\partial_Z} \).

(iii) \( Y = \alpha'(\hat{x})[\text{cone}(\Omega - \{\hat{x}\})] \cap \text{ker}(\beta'(\hat{x})) \) and \( Z = \beta'(\hat{x})[\text{cone}(\Omega - \{\hat{x}\})] \).

If \( K \) and \( \Omega \) both have a nonempty interior then (i) to (iii) are equivalent to:

(iv) \exists \alpha \in \text{int}(\Omega): \alpha(\hat{x}) + \alpha'(\hat{x})[x - \hat{x}] = \text{int}(K) \land \beta'(\hat{x})[x - \hat{x}] = \partial_Z \land \beta'(\hat{x}) \text{ is surjective.}

The constraint qualification (iv) of Lemma 2.1 is nothing else but the popular Mangasarian-Fromovitz-Constraint-Qualification MFCQ, which is well-known from finite-dimensional optimization. However, due to the fact that many of the common ordering cones in infinite-dimensional spaces have an empty interior, MFCQ can not be applied to problem (3) in several cases. That is why we often have to deal with KRZCQ.

Let \( \hat{x} \in M \) be an arbitrarily chosen feasible point of problem (3). Then we define the linearized tangent cone to \( M \) at \( \hat{x} \) as follows:

\[
\mathcal{L}_M(\hat{x}) := \{ d \in \text{cl}(\text{cone}(\Omega - \{\hat{x}\})) \mid \alpha'(\hat{x})[d] \in - \text{cl}(\text{cone}(K + [\alpha(\hat{x})])), \beta'(\hat{x})[d] = \partial_Z \}.
\]

One can easily check that \( \mathcal{L}_M(\hat{x}) \) is a closed, convex cone satisfying \( \mathcal{T}_M(\hat{x}) \subseteq \mathcal{L}_M(\hat{x}) \) where \( \mathcal{T}_M(\hat{x}) \) is the Bouligand tangent cone to \( M \) at \( \hat{x} \) defined as follows:

\[
\mathcal{T}_M(\hat{x}) := \{ d \in X \mid \exists d_k \subseteq X \exists t_k \subseteq \mathbb{R}^+ : \lim_{k \to \infty} d_k = d \land \lim_{k \to \infty} t_k = 0 \land \hat{x} + t_k d_k \in M \forall k \in \mathbb{N} \}.
\]

In order to ensure that the opposite inclusion holds true as well, one has to postulate some constraint qualification to hold.

**Lemma 2.2.** [5]

Let \( \hat{x} \in M \) be a point where KRZCQ holds. Then we have \( \mathcal{T}_M(\hat{x}) = \mathcal{L}_M(\hat{x}) \).

Using the lemma above we can derive a necessary optimality condition of KKT-type for problem (3).

**Theorem 2.1.**

Let \( \hat{x} \in M \) be a local optimal solution of (3) where KRZCQ holds. Then there exist a subgradient \( \xi \in \partial \psi(\hat{x}) \), \( \lambda \in K^0 \) and \( \mu \in Z' \) such that the following conditions are satisfied:

\[
\forall d \in \text{cl}(\text{cone}(\Omega - \{\hat{x}\})) : \quad (\xi + \alpha'(\hat{x}) \circ \lambda + \beta'(\hat{x}) \circ \mu)[d] \geq 0, \tag{5}
\]

\[
\langle \alpha(\hat{x}), \lambda \rangle = 0. \tag{6}
\]
Proof:
Since $\psi$ is a locally Lipschitz continuous functional, it is directionally differentiable in the sense of Clarke. Moreover, the fact that $\hat{x} \in M$ is a local minimizer of (3) leads to (adaption of Theorem 5.3.1 in [3]):

$$\forall d \in \mathcal{T}_M(\hat{x}): \psi^*_f(\hat{x}; d) \geq 0.$$  

(7)

Since KRZCQ holds at $\hat{x}$, we deduce $\mathcal{T}_M(\hat{x}) = \mathcal{L}_M(\hat{x})$ from Lemma 2.2. Considering the set

$$S := \{(\psi^*_f(\hat{x}; d) + v, \alpha'(\hat{x})[d] + w, \beta'(\hat{x})[d]) \in \mathbb{R} \times Y \times Z \mid d \in \text{cl}(\text{cone}(\Omega - \{\hat{x}\})), v \geq 0, w \in \text{cl}(\text{cone}(K + \{\alpha(\hat{x})\}))\}$$

we easily see from (7) that $(0, \sigma_Y, \sigma_Z)$ is a boundary point of $S$. Since $S$ is a convex set with nonempty interior (cf. statement (ii) of Lemma 2.1), we can use Theorem C.2 in [13] to separate $(0, \sigma_Y, \sigma_Z)$ from $S$ and derive the existence of $\lambda_0 \in \mathbb{R}, \lambda \in \mathcal{Y}^*$ and $\mu \in \mathcal{Z}^*$ satisfying $(\lambda_0, \lambda, \mu) \neq (0, \sigma_Y, \sigma_Z)$ and

$$\forall d \in \text{cl}(\text{cone}(\Omega - \{\hat{x}\})) \forall v \in \mathbb{R}_+^n \forall w \in \text{cl}(\text{cone}(K + \{\alpha(\hat{x})\})):\n\lambda_0(\psi^*_f(\hat{x}; d) + v) + (\alpha'(\hat{x})[d] + w, \lambda) + (\beta'(\hat{x})[d], \mu) \geq 0.$$  

Similar to the proof of Theorem 5.3 in [13] we obtain $\lambda_0 \geq 0, \lambda \in \mathcal{K}^D$ and (6) as well as:

$$\forall d \in \text{cl}(\text{cone}(\Omega - \{\hat{x}\})) : (\lambda_0 \cdot \psi^*_f(\hat{x}; \cdot) + \alpha'(\hat{x}) \circ \lambda + \beta'(\hat{x}) \circ \mu)[d] \geq 0.$$  

(8)

Suppose that we have $\lambda_0 = 0$. Since KRZCQ holds, for any $y \in \mathcal{Y}$ there is $d \in \text{cone}(\Omega - \{\hat{x}\})$ satisfying $\beta'(\hat{x})[d] = \sigma_Z$ and some $w \in \text{cone}(K + \{\alpha(\hat{x})\})$ such that $y = \alpha'(\hat{x})[d] + w$ holds true (statement (iii) of Lemma 2.1). From (6), (8) and $\lambda \in \mathcal{K}^D$ we get:

$$(y, \lambda) = (\alpha'(\hat{x})[d] + w, \lambda) = (\alpha'(\hat{x}) \circ \lambda)[d] + (w, \lambda) = (\alpha'(\hat{x}) \circ \lambda + \beta'(\hat{x}) \circ \mu)[d] + (w, \lambda) \geq 0.$$  

Since $y$ was chosen arbitrarily, we have $\lambda = \sigma_Y$. Moreover, we have $\beta'(\hat{x})[\text{cone}(\Omega - \{\hat{x}\})] = \mathcal{Z}$ from KRZCQ (statement (iii) of Lemma 2.1) so using (8) we obtain $\mu[Z] = (\beta'(\hat{x}) \circ \mu)[\text{cone}(\Omega - \{\hat{x}\})] \subseteq \mathbb{R}_+^n$ and hence $\mu = \sigma_Z$ since $\mu[Z]$ is a subspace of $\mathbb{R}$. Consequently, we have $(\lambda_0, \lambda, \mu) = (0, \sigma_Y, \sigma_Z)$ which is a contradiction to the separation theorem used above. As a result, we can assume now w.l.o.g. $\lambda_0 = 1$.

Now we have from (8) with $\lambda_0 = 1$ that $\hat{d} = \sigma_X$ is an optimal solution of the convex optimization problem

$$\min_{d \in \Omega(\hat{x})} \{\psi^*_f(\hat{x}; \cdot) + \alpha'(\hat{x}) \circ \lambda + \beta'(\hat{x}) \circ \mu)[d] \mid d \in \Omega(\hat{x})\}$$

where $\Omega(\hat{x}) := \text{cl}(\text{cone}(\Omega - \{\hat{x}\}))$. As a consequence, we have

$$\sigma_X \in \partial(\psi^*_f(\hat{x}; \cdot) + \alpha'(\hat{x}) \circ \lambda + \beta'(\hat{x}) \circ \mu)(\sigma_X) + \mathcal{N}_{\Omega(\hat{x})}(\sigma_X) = \partial \psi^*_f(\hat{x}) + \{\alpha'(\hat{x}) \circ \lambda + \beta'(\hat{x}) \circ \mu\} - (\Omega(\hat{x}))^D$$

where $\partial(\cdot)$ denotes the subdifferential of convex analysis. That is why we find some $\xi \in \partial \psi^*_f(\hat{x})$ and some $\varrho \in -(\Omega(\hat{x}))^D$ such that $\xi + \alpha'(\hat{x}) \circ \lambda + \beta'(\hat{x}) \circ \mu = -\varrho$ holds. Hence, for arbitrary $d \in \text{cl}(\text{cone}(\Omega - \{\hat{x}\}))$ we have from the definition of $\Omega(\hat{x})$

$$(\xi + \alpha'(\hat{x}) \circ \lambda + \beta'(\hat{x}) \circ \mu)[d] = (\varrho)[d] \geq 0$$

which equivalently means that $\xi, \lambda$ and $\mu$ also satisfy condition (5).

3. Partial calmness

Bilevel programming problems are hierarchical optimization problems with two decision levels appearing in numerous applications. In the so-called upper level a leader has to fix a variable which is given to the follower, represented by
the so-called lower level, who has to solve an optimization problem now which depends on the leader’s variable (i.e. a parametric optimization problem where the parameter is the variable of the leader). After passing a solution to the upper level the leader is able to evaluate his objective function which does not only depend on his variable but also on the follower’s one. Obviously, this problem is not well-defined whenever there exist values for the leader’s variable where the corresponding solution set of the follower is not a singleton. In this case one can distinguish between several approaches to interpret the problem. The optimistic approach models a cooperative behavior of leader and follower while the pessimistic one depicts a competitive situation (cf. [7]). Furthermore, the whole problem may be interpreted as a problem of set optimization where a set-valued objective functional has to be optimized in a particular sense (cf. [17]). Since all the different approaches result in completely different optimality conditions, bilevel programming is a wide field of research which is still very popular although the first bilevel programming problem has already been introduced by Stackelberg in 1934 (cf. [19]).

The given bilevel programming problem (1), (2) models a cooperative situation which is similar to the optimistic approach. In order to apply the necessary optimality conditions of Theorem 2.1 to the bilevel optimization problem (1), (2) one has to find a single level reformulation of this problem. Hence, let us define two set-valued mappings $\Gamma, \Psi : \mathbb{R}^n \longrightarrow \mathcal{P} (\mathbb{R}^k)$ and an extended real-valued function $\varphi : \mathbb{R}^c \longrightarrow \mathbb{R}$ as follows:

$$\forall z \in \mathbb{R}^n : \quad \Gamma (z) := \left\{ y \in \mathbb{R}^k \mid g(z,y) \leq \sigma_{Gr}, h(z,y) = \sigma_{Gr} \right\}$$

$$\varphi(z) := \begin{cases} +\infty & \text{if } \Gamma(z) = \emptyset \\ \inf \{f(z,y) \mid y \in \Gamma(z)\} & \text{if } \Gamma(z) \neq \emptyset \\ \emptyset & \text{if } \Gamma(z) = \emptyset \end{cases}$$

$$\Psi(z) := \left\{ y \in \Gamma(z) \mid f(z,y) = \varphi(z) \right\}.$$

In the context of the parametric optimization problem (2) we call $\Gamma$ and $\Psi$ feasible set mapping and solution set mapping, respectively. Furthermore, $\varphi$ is called optimal value function of (2) and is likely to be nonsmooth.

We are going to consider the optimal value reformulation of problem (1), (2) which is given as follows:

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

$$\dot{x}(t) = F(t,x(t),u(t)) \quad \text{a.e. } t \in [0,T]$$

$$x(0) = x_0 \quad \text{a.e. } t \in [0,T]$$

$$G(t,u(t)) \leq \sigma_{Gr}$$

$$H(x(T)) = \sigma_{Gr}$$

$$f(x(T),y) - \varphi(x(T)) \leq 0$$

$$g(x(T),y) \leq \sigma_{Gr}$$

$$h(x(T),y) = \sigma_{Gr}.$$

The idea to replace the lower level problem (2) by inserting the constraints $f(x(T),y) - \varphi(x(T)) \leq 0$, $g(x(T),y) \leq \sigma_{Gr}$ and $h(z,y) = \sigma_{Gr}$ into (1) was introduced by Ye and Zhu in [21] in the finite-dimensional case. In the latter paper the authors make clear that the new problem they derived is equivalent to the original one (refer also to Theorem 3.1 in [10]) but common constraint qualifications like KRZCQ, MFCQ and LICQ (or at least their nonsmooth counterparts) fail at any feasible point of the reformulation. That is why they use the concept of partial calmness in order to shift the crucial constraint involving the optimal value function of the lower level problem to the objective function. The emerging problem can be interpreted as a problem of type (3) provided $\varphi$ is locally Lipschitz continuous and this property can be ensured (at least in a neighbourhood of a reference point) if common constraint qualifications (cf. [7, 9]) hold at this point w.r.t. the lower level problem (2). There are several publications (e.g. [7, 8, 10]) that give necessary optimality conditions for optimistic bilevel programming problems under partial calmness in finite-dimensional spaces. In [20], the author investigates general optimistic bilevel optimal control problems and gives an abstract definition of partial calmness in infinite-dimensional spaces. In this section we will give the definition of partial calmness according to problem (1), (2). Furthermore, we provide some criteria that ensure that partial calmness
is given automatically.
Before we start our considerations let us define a functional $Z : E^2 \rightarrow \mathbb{R}$ by:

$$
\forall (x, u, y) \in E^2 : Z(x, u, y) := F_0(x(T), y) + \int_0^T F_1(t, x(t), u(t), y) dt.
$$

Note that $Z$ equals the objective functional of the upper level problem (1). Furthermore, $Z$ is continuously Fréchet differentiable since $F_0$ and $F_1$ are assumed to be continuously differentiable, $x$ comes from the space $W^{1,\infty}_{loc}[0, T]$ and $u$ is an element of $L^\infty_{loc}[0, T]$.

**Definition 3.1. (Local optimal solution)**

A feasible point $(\hat{x}, \hat{u}, \hat{y}) \in E^2$ of problem (1), (2) is called local minimizer or local optimal solution if there exists $\varepsilon > 0$ such that for any feasible point $(x, u, y) \in U^\varepsilon_E(\hat{x}, \hat{u}, \hat{y})$ of (1), (2) we have:

$$
Z(\hat{x}, \hat{u}, \hat{y}) \leq Z(x, u, y).
$$

In the theory of optimal control the above definition of a local optimal solution corresponds with the concept of weak local minimizers (cf. [11]) since we are respecting the norm of the control function $u$ as well.

In order to define the concept of partial calmness one has to consider the following partially perturbed problem:

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

$$
\begin{align*}
\dot{x}(t) &= F(t, x(t), u(t)) \quad \text{a.e. } t \in [0, T] \\
x(0) &= x_0 \\
G(t, u(t)) &\leq \sigma_{G^c} \quad \text{a.e. } t \in [0, T] \\
H(x(T)) &= \sigma_{G^c} \\
f(x(T), y) - \varphi(x(T)) &\leq \kappa \\
g(x(T), y) &\leq \sigma_{G^c} \\
h(x(T), y) &= \sigma_{G^c}.
\end{align*}
$$

(10)

**Definition 3.2. (Partial calmness)**

Let $(\hat{x}, \hat{u}, \hat{y}) \in E^2$ be a local optimal solution of problem (1), (2). Problem (9) is called partially calm at $(\hat{x}, \hat{u}, \hat{y})$ if there exist constants $\delta, \eta > 0$ such that for any feasible point $(x, u, y, \kappa) \in U^\delta_{E^3}(\hat{x}, \hat{u}, \hat{y}, 0)$ of (10) we have:

$$
Z(x, u, y) - Z(\hat{x}, \hat{u}, \hat{y}) + \eta \kappa \geq 0.
$$

(11)

Now we are able to show a basic characterization of partial calmness which is similar to a result in [21].
Lemma 3.1.
Let \((\hat{x}, \hat{u}, \hat{y}) \in \mathcal{E}_3 \) be a local optimal solution of (1), (2). Then (9) is partially calm at \((\hat{x}, \hat{u}, \hat{y})\) if and only if there exists a constant \(\rho > 0\) such that \((\hat{x}, \hat{u}, \hat{y})\) is a local optimal solution of:

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

\[
\begin{align*}
\dot{x}(t) &= F(t, x(t), u(t)) & \text{a.e. } t \in [0, T] \\
x(0) &= x_0 \\
G(t, u(t)) &\leq \varphi_{\rho'} & \text{a.e. } t \in [0, T] \\
H(x(T)) &= \varphi_{\rho'} \\
g(x(T), y) &\leq \varphi_{\rho'} \\
h(x(T), y) &= \varphi_{\rho'}.
\end{align*}
\]

Proof:
[\(\Rightarrow\)] Since (9) is partially calm at \((\hat{x}, \hat{u}, \hat{y})\), we find \(\varepsilon > 0\) and \(\eta > 0\) such that for any point \((x, u, y, \kappa) \in \bigcup_{\mathcal{E}_3}^T (\hat{x}, \hat{u}, \hat{y}, 0)\), feasible for (10), we have (11).

Now choose an arbitrary point \((x, u, y) \in \bigcup_{\mathcal{E}_3}^T (\hat{x}, \hat{u}, \hat{y})\) which is feasible for problem (12). We consider two different cases.

Case 1: Let \(\kappa := f(x(T), y) - \varphi(x(T)) < \varepsilon\) hold. Since we have \(\kappa \geq 0\) from \(f(x(T), y) - \varphi(x(T)) \geq 0\), we conclude \((x, u, y, \kappa) \in \bigcup_{\mathcal{E}_3}^T (\hat{x}, \hat{u}, \hat{y}, 0)\) and this point is feasible for (10). Hence, we deduce from the partial calmness:

\[
\eta(f(\hat{x}(T), \hat{y}) - \varphi(\hat{x}(T))) + Z(\hat{x}, \hat{u}, \hat{y}) = Z(\hat{x}, \hat{u}, \hat{y}) \\
\leq \eta \kappa + Z(x, u, y) = \eta(f(x(T), y) - \varphi(x(T))) + Z(x, u, y).
\]

Case 2: Let \(f(x(T), y) - \varphi(x(T)) \geq \varepsilon\) hold. Since \(Z\) is continuously Fréchet differentiable, it is Lipschitz continuous on the set \(\bigcup_{\mathcal{E}_3}^T (\hat{x}, \hat{u}, \hat{y})\) with Lipschitz constant \(L_Z > 0\). That is why we have:

\[
Z(\hat{x}, \hat{u}, \hat{y}) \leq Z(x, u, y) + L_Z \| (x, u, y) - (\hat{x}, \hat{u}, \hat{y}) \|_{\mathcal{E}_3} < Z(x, u, y) + L_Z \varepsilon.
\]

Hence, we conclude:

\[
L_Z(f(\hat{x}(T), \hat{y}) - \varphi(\hat{x}(T))) + Z(\hat{x}, \hat{u}, \hat{y}) = Z(\hat{x}, \hat{u}, \hat{y}) \\
< Z(x, u, y) + L_Z \varepsilon \leq L_Z(f(x(T), y) - \varphi(x(T))) + Z(x, u, y).
\]

Combining (13) and (14) we have shown that \((\hat{x}, \hat{u}, \hat{y})\) is a local optimal solution of (12) for \(\rho = \max\{\eta; L_Z\}\) (in fact for any \(\rho \geq \max\{\eta; L_Z\}\)).

[\(\Leftarrow\)] Let \((\hat{x}, \hat{u}, \hat{y})\) be a local optimal solution of problem (12). Then there exists \(\varepsilon > 0\) such that for any point \((x, u, y, \kappa) \in \bigcup_{\mathcal{E}_3}^T (\hat{x}, \hat{u}, \hat{y}, 0)\), feasible for (10), we have:

\[
\rho(f(x(T), y) - \varphi(x(T))) + Z(x, u, y) \geq \rho(f(\hat{x}(T), \hat{y}) - \varphi(\hat{x}(T))) + Z(\hat{x}, \hat{u}, \hat{y}).
\]

Since \((\hat{x}, \hat{u}, \hat{y})\) is feasible for (9), we have \(f(\hat{x}(T), \hat{y}) - \varphi(\hat{x}(T)) = 0\) and hence:

\[
Z(x, u, y) - Z(\hat{x}, \hat{u}, \hat{y}) + \rho \kappa \geq Z(x, u, y) - Z(\hat{x}, \hat{u}, \hat{y}) + \rho(f(x(T), y) - \varphi(x(T))) \geq 0.
\]

That is why (9) is partially calm at \((\hat{x}, \hat{u}, \hat{y})\) with \(\eta := \rho\). 

Now we investigate some conditions that ensure partial calmness of problem (9) at a local optimal solution \((\hat{x}, \hat{u}, \hat{y})\). In [12] Henrion and Surowiec introduce the so-called value-function-constraint-qualification VFCQ for finite-dimensional
bilevel programming problems which is likely to guarantee partial calmness. However, VFCQ is too weak for problems of type (1), (2) since the lower level problem just deals with the final state of the state function \( x \). That is why we give an adaption of VFCQ below, which fits problem (1), (2). Therefore, we will consider a special set-valued mapping \( \Delta : \mathbb{R} \rightarrow \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^k) \) defined as follows:

\[
\forall \kappa \in \mathbb{R} : \quad \Delta(\kappa) = \{(z, y) \in \text{graph}(\Gamma) \mid f(z, y) - \varphi(z) \leq \kappa \}. \tag{15}
\]

Note that \( \Delta(\kappa) = \emptyset \) holds whenever \( \kappa < 0 \). Furthermore, \( \Delta(0) = \text{graph}(\Psi) \) and \( \text{graph}(\Psi) \subseteq \Delta(\kappa) \) for any \( \kappa \geq 0 \) are obvious from the definition of \( \Delta \).

**Definition 3.3.** (EVFCQ)

Let \((\hat{x}, \hat{u}, \hat{y}) \in E_2\) be a feasible point of the bilevel programming problem (1), (2). Then EVFCQ, the enhanced-value

function-constraint-qualification, holds at \((\hat{x}, \hat{u}, \hat{y})\) provided there exist constants \( \delta_1, \delta_2, L > 0 \) such that the condition

\[
\forall \kappa \in [0, \delta_1) : \quad \Delta(\kappa) \cap \mathcal{U}_{\mathbb{R}^n \times \mathbb{R}^k}^{\delta_2}((\hat{x}(T), \hat{y})) \subseteq \text{graph}(\Psi) + \{0_{\mathbb{R}^n} \times \kappa \cdot \mathbb{B}_{\mathbb{R}^k}
\]

holds.

Note that if EVFCQ holds at a feasible point \((\hat{x}, \hat{u}, \hat{y})\) of (1), (2), then \( \Delta \) is calm at \((0, \hat{x}(T), \hat{y})\) in the sense of set-valued mappings. The converse is not true in general (which is the difference to the original value-function-constraint-qualification due to Henrion and Surowiec; the latter one just says that \( \Delta \) is calm at \((0, \hat{x}(T), \hat{y})\)). Furthermore, EVFCQ holds at a feasible point \((\hat{x}, \hat{u}, \hat{y})\) of (1), (2) if and only if there are constants \( \delta_1, \delta_2, L > 0 \) such that we have:

\[
\forall \kappa \in [0, \delta_1) \forall (z, y) \in \Delta(\kappa) \cap \mathcal{U}_{\mathbb{R}^n \times \mathbb{R}^k}^{\delta_2}((\hat{x}(T), \hat{y})) \exists \Psi(z) : \quad \|y - \Psi(z)\|_\infty \leq L \cdot \kappa. \tag{16}
\]

**Lemma 3.2.**

Let \((\hat{x}, \hat{u}, \hat{y}) \in E_2\) be a local optimal solution of the bilevel optimal control problem (1), (2) where EVFCQ holds. Then (9) is partially calm at \((\hat{x}, \hat{u}, \hat{y})\).

**Proof:**

Since \((\hat{x}, \hat{u}, \hat{y})\) is also a local optimal solution of (9), we find \( \varepsilon > 0 \) such that for any feasible point \((x, u, y) \in \mathcal{U}_{E_2}^\varepsilon((\hat{x}, \hat{u}, \hat{y})\) of (9) we have \( Z(x, u, y) \geq Z(\hat{x}, \hat{u}, \hat{y}) \). Furthermore, from EVFCQ we deduce the existence of constants \( \delta_1, \delta_2, L > 0 \) so that (16) holds.

Now we choose an arbitrary point \((x, u, y, \kappa) \in \mathcal{U}_{E_2}^\varepsilon((\hat{x}, \hat{u}, \hat{y})\) which is feasible for (10) where the constant \( \delta > 0 \) is defined as\( \delta := \min \{\varepsilon; \delta_1; \delta_2; \frac{1}{1+L}\} \). From the definition of \( \Delta \) in (15) we have \((x, y) \in \Delta(\kappa)\) and from \( \|x - \hat{x}\|_{\mathbb{R}^n, [0, T]} < \delta \) we deduce \( \|x(T) - \hat{x}(T)\|_\infty < \delta \leq \delta_2 \) which gives us \((x(T), y) \in \mathcal{U}_{\mathbb{R}^n \times \mathbb{R}^k}^{\delta_2}((\hat{x}(T), \hat{y}))\). Consequently, from (16) we derive the existence of a vector \( \bar{y} \in \Psi(x(T)) \) such that

\[
\|y - \bar{y}\|_\infty \leq L \cdot \kappa
\]

is true. Moreover, \( \bar{y} \) satisfies

\[
\|\bar{y} - \bar{y}\|_\infty \leq \|y - \bar{y}\|_\infty + \|y - \bar{y}\|_\infty < \delta + L \cdot \kappa < (1 + L)\delta \leq (1 + L) \cdot \frac{\varepsilon}{1+L} = \varepsilon
\]

so that we have \((x, u, \bar{y}) \in \mathcal{U}_{E_2}^\varepsilon((\hat{x}, \hat{u}, \hat{y})\). Since \((x, u, \bar{y})\) is feasible for (9) (due to \( \bar{y} \in \Psi(x(T)) \)), we deduce \( Z(x, u, \bar{y}) \geq Z(\hat{x}, \hat{u}, \hat{y}) \).

Observe that \( Z \) is continuously Fréchet differentiable and hence locally Lipschitz continuous at \((\hat{x}, \hat{u}, \hat{y})\). Let \( L_Z \) denote its Lipschitz constant on the set \( \mathcal{U}_{E_2}^\varepsilon((\hat{x}, \hat{u}, \hat{y})\). Since \((x, u, y)\) is an element of this set, we finally have:

\[
Z(\hat{x}, \hat{u}, \hat{y}) - Z(x, u, y) \leq Z(\hat{x}, \hat{u}, \hat{y}) - Z(x, u, \bar{y}) \leq L_Z \cdot \|y - \bar{y}\|_\infty \leq L_Z \cdot L \cdot \kappa.
\]

Now we rearrange the inequality above to derive:

\[
Z(x, u, y) - Z(\hat{x}, \hat{u}, \hat{y}) + L_Z \cdot L \cdot \kappa \geq 0.
\]
Consequently, (9) is partially calm at \((\hat{x}, \hat{u}, \hat{y})\) with \(\eta := L_Z \cdot L\).  

Another well-known condition that ensures partial calmness is the property of (2) to possess a so-called uniformly weak sharp minimum. This constraint qualification has originally been introduced by Ye and Zhu in [21]. In the latter paper the authors give several conditions for (2) to possess a uniformly weak sharp minimum. We are going to adapt their results here. The only question arising is whether the existence of a uniformly weak sharp minimum for (2) implies partial calmness of (9) since the upper level problem (1) is infinite-dimensional. Before we deal with this problem we formulate the definition of uniformly weak sharp minima.

**Definition 3.4. (Uniformly weak sharp minimum)**

The parametric optimization problem (2) is said to possess a uniformly weak sharp minimum provided there exists a constant \(\gamma > 0\) such that we have:

\[
\forall (z, y) \in \text{graph}(\Gamma): \quad f(z, y) - \varphi(z) \geq \gamma \cdot \min_{w \in \Psi(z)} \|w - y\|_\infty.
\]

Note that the existence of a uniformly weak sharp minimum implies that EVFCQ holds at any local optimal solution of problem (1), (2) under one technical assumption.

**Lemma 3.3.**

Let (2) possess a uniformly weak sharp minimum. Then EVFCQ holds at any local optimal solution \((\hat{x}, \hat{u}, \hat{y}) \in E_2\) of (1), (2) provided \(\hat{x}(T) \in \text{int}(\text{dom}(\Psi))\) is satisfied.

**Proof:**

Let \(\delta_2 > 0\) be chosen in such a way that we have \(\cup_{z \in \Delta(\hat{x})}^\delta(\hat{x}(T)) \subseteq \text{dom}(\Psi)\). Moreover, fix an arbitrary \(0 \leq \kappa \leq \delta_1\) where \(\delta_1\) is an arbitrary positive real number and choose \((z, y) \in \Delta(\kappa) \cap \cup_{z \in \Delta(\hat{x})}^\delta(\hat{x}(T), \hat{y})\) (note that \((\hat{x}(T), \hat{y}) \in \Delta(\kappa))\). From the definition of a uniformly weak sharp minimum and \((z, y) \in \text{graph}(\Gamma)\) we have the existence of a constant \(\gamma > 0\) such that the following holds true:

\[
\min_{w \in \Psi(z)} \|w - y\|_\infty \leq \frac{\gamma}{\gamma} [f(z, y) - \varphi(z)] \leq \frac{\gamma}{\gamma}.
\]

Since we have \(z \in \text{dom}(\Psi)\), there is \(\check{y} \in \Psi(z)\) satisfying \(\|\check{y} - y\|_\infty \leq \frac{\gamma}{\gamma}\). Hence, EVFCQ holds at \((\hat{x}, \hat{u}, \hat{y})\) with \(L = \frac{1}{\gamma}\).

Although our version of VFCQ is slightly stronger than the one introduced by Henrion and Surowiec, it is weaker than the existence of a uniformly weak sharp minimum of (2). We just mention a little example here.

**Example 3.1. [12]**

Consider the parametric optimization problem (2) given as follows:

\[
\min_y \{ zy | 0 \leq y \leq 2 \}.
\]

Think of a feasible point \((\hat{x}, \hat{u}, \hat{y})\) of some problem (1), (2) where \(\hat{x}(T) = \hat{z} = -1\) and \(\hat{y} = 2\) hold. Then EVFCQ is satisfied at this point although (2) does not possess a uniformly weak sharp minimum.

One easily computes:

\[
\forall z \in \mathbb{R}: \quad \Psi(z) = \begin{cases} 
2 & \text{if } z < 0 \\
[0, 2] & \text{if } z = 0 \\
[0] & \text{if } z > 0
\end{cases} \quad \varphi(z) = \begin{cases} 
2z & \text{if } z < 0 \\
0 & \text{if } z \geq 0.
\end{cases}
\]
Now take $\delta_1 = \delta_2 = \frac{1}{2}$ and $L = 2$. We consider $(z, y) \in \Delta(\kappa) \cap \mathcal{U}_{\mathbb{R}^2}((\bar{z}, \bar{y})$). From the definition of $\Delta$ we derive $z(y - 2) \leq \kappa$. Besides, we have $z \in \left(-\frac{1}{2}, -\frac{1}{2}\right)$ and $y \in \left[\frac{1}{2}, 2\right]$. As a consequence, $y \in \left[\max\left\{\frac{1}{2}, 2 + \frac{1}{z}\right\}, 2\right]$ holds true and we can conclude from $\Psi(z) = \{2\}$:

$$\min_{w \in \Psi(z)} |w - y| = |2 - y| \leq \left|2 - 2 - \frac{\kappa}{z}\right| = \left|\frac{\kappa}{z}\right| < 2\kappa.$$  
Hence, EVFCQ holds at $(\bar{z}, \bar{u}, \bar{y})$.

Be aware of the fact that the given parametric optimization problem does not possess a uniformly weak sharp minimum since:

$$\forall (z, 0) \in \text{graph}(\Gamma): \quad z < 0 \implies f(z, 0) - \varphi(z) = -2z \wedge \min_{w \in \Psi(z)} |w - 0| = 2.$$  
That means if we consider the sequence $\left(-\frac{1}{2}, 0\right) \subseteq \text{graph}(\Gamma)$, then $f\left((-\frac{1}{2}, 0)\right) - \varphi\left(-\frac{1}{2}\right) = \frac{\kappa}{2}$ tends to zero and consequently the constant $y$ in Definition 3.4 of the uniformly weak sharp minimum has to be zero. However, this choice was excluded.  
In a similar way one can check that EVFCQ fails to hold at some point $(\bar{x}, \bar{u}, \bar{y}) \in E_2$ satisfying $\bar{x}(T) = \bar{z} = 0$ and $\bar{y} = 1$: For arbitrary $\kappa > 0$ and $0 < \delta_2 < 1$ the point $(z, y) = \left(\max\left\{-\frac{\kappa}{2}, \frac{\kappa}{2} - \frac{\kappa}{2}\right\}, 1 + \frac{\kappa}{2}\right)$ belongs to $\Delta(\kappa) \cap \mathcal{U}_{\mathbb{R}^2}((\bar{z}, \bar{y})$. Since $\Psi(z) = \{2\}$ holds, we have $\min_{w \in \Psi(z)} |w - y| = \left|2 - 1 - \frac{\kappa}{2}\right| = \frac{\kappa}{2}$, but this value does not depend linearly on $\kappa$. This means that the constant $L$ from the definition of EVFCQ does not exist even for arbitrarily small $\delta_1$ and $\delta_2$.  
Observe that the original regularity condition VFCQ due to Henrion and Surowiec holds at $(\bar{z}, \bar{y})$ (cf. [12]) which shows that EVFCQ is indeed stronger than VFCQ.  

From Lemma 3.2 and Lemma 3.3 we easily deduce the following remark.

**Remark 3.1.**  
Let (2) possess a uniformly weak sharp minimum. If $(\bar{x}, \bar{u}, \bar{y}) \in E_2$ is a local optimal solution of (1), (2) such that $\bar{x}(T) \in \text{int}(\text{dom}(\Psi))$ holds true, then (9) is partially calm at $(\bar{x}, \bar{u}, \bar{y})$.

In [21] one can find several conditions which guarantee that (2) possesses a uniformly weak sharp minimum. We mention two interesting results here.

**Lemma 3.4.** [21]  
Consider matrices $A_1 \in \mathbb{R}^{l \times n}, A_2 \in \mathbb{R}^{l \times k}, Q \in \mathbb{R}^{k \times k}, a_1 \in \mathbb{R}^k$ as well as $a_2 \in \mathbb{R}^l$ where $Q$ is positive semidefinite and symmetric.

(i) The parametric optimization problem

$$\min_{y} \left\{ a_1^T y \mid A_1 z + A_2 y \leq a_2 \right\}$$  
(17)

possesses a uniformly weak sharp minimum.

(ii) Consider the following quadratic optimization problem:

$$\min_{y} \left\{ y^T Q y + a_1^T y \mid A_1 z + A_2 y \leq a_2 \right\}.$$  
(18)

If either

$$\forall (z, y) \in \text{graph}(\Psi): \quad Q y + a_1 = a_{2l}$$  

or

$$\exists C > 0 \forall (z, y) \in \text{graph}(\Psi): \quad Q y + a_1 \neq a_{2l} \implies \|Q y + a_1\|_{\infty} \geq C$$

holds, then (18) possesses a uniformly weak sharp minimum provided the following condition is satisfied:

$$\forall (z, y) \in \text{graph}(\Psi) \exists \bar{y} \in \Psi(z): \quad \left| d \in \mathbb{R}^l \mid d^T (Q \bar{y} + a_1) = 0, d \in T_{\Psi(z)}(y) \right| \subseteq \ker(Q).$$
4. Linearized Pontryagin Maximum Principle

In order to derive a necessary optimality condition of Pontryagin-type for problem (1), (2) we first have to point out some basic facts from functional analysis. Of essential importance in this section will be the matrix function \( \Phi \) which is the uniquely determined solution of the matrix differential equation

\[
\Phi(t) = \nabla_x F(t, \hat{x}(t), \hat{u}(t)) \Phi(t) \quad \text{a.e. } t \in [0, T] \quad \Phi(0) = I_n
\]

(19)

where \((\hat{x}, \hat{u}, \hat{y}) \in E_2\) is a feasible point for problem (1), (2).

The following lemma summarizes some basic results on a special Volterra equation of second kind. The proof is omitted here; one may find it part of the proof of Theorem 5.19 in [13].

**Lemma 4.1.**

Let \( \omega \in W_{1,\infty}^n[0, T] \) and a feasible point \((\hat{x}, \hat{u}, \hat{y}) \in E_2\) of (1), (2) be fixed. Considering the Volterra-equation of second kind

\[
\omega(\cdot) = x(\cdot) - \int_0^\cdot \nabla_x F(\tau, \hat{x}(\tau), \hat{u}(\tau))x(\tau) d\tau,
\]

(20)

the following two statements hold true.

(i) Equation (20) has a (not necessarily unique) solution.

(ii) One of the solutions of (20) is given as follows:

\[
\forall t \in [0, T]: \quad x(t) := \omega(t) + \Phi(t) \int_0^t \Phi^{-1}(\tau) \nabla_x F(\tau, \hat{x}(\tau), \hat{u}(\tau)) \omega(\tau) d\tau.
\]

(21)

As said earlier the partial calmness property of (9) allows us to transform (1), (2) locally into the problem (12). If we additionally could ensure the local Lipschitz continuity of \( \varphi \) then the latter problem is of type (3) and we could apply Theorem 2.1 in order to find necessary optimality conditions for (12) and hence for (1), (2). However, in that case we need to know more about the Clarke subdifferential of \( \varphi \). Since we want to omit a precise calculation of the lower level optimal value function or its Clarke subdifferential, we are going to use an approximation formula.

**Lemma 4.2.** [9]

Let \((\tilde{z}, \tilde{y}) \in \text{graph}(\Psi)\) be a feasible point of (2) where MFCQ w.r.t. \( y \) holds, i.e. \( \nabla_y h(\tilde{z}, \tilde{y}) \) has rank \( q \) and there exists \( d_y \in \mathbb{R}^q \) such that

\[
g(\tilde{z}, \tilde{y}) + \nabla_y g(\tilde{z}, \tilde{y}) d_y < 0, \quad \nabla_y h(\tilde{z}, \tilde{y}) d_y = 0.
\]

is satisfied. If \( \Psi \) is inner semicontinuous at \((\tilde{z}, \tilde{y})\), then \( \varphi \) is locally Lipschitz continuous at \( \tilde{z} \) and we have

\[
\partial^c \varphi(\tilde{z}) \subseteq \left\{ \nabla_x f(\tilde{z}, \tilde{y}) + \lambda^T \nabla_y g(\tilde{z}, \tilde{y}) + \mu^T \nabla_y h(\tilde{z}, \tilde{y}) \mid (\lambda, \mu) \in \Lambda(\tilde{z}, \tilde{y}) \right\}
\]

where \( \Lambda(\tilde{z}, \tilde{y}) \) is the set of regular Lagrange multipliers of (2) at \((\tilde{z}, \tilde{y})\) defined as follows:

\[
\Lambda(\tilde{z}, \tilde{y}) := \left\{ (\lambda, \mu) \in \mathbb{R}^p \times \mathbb{R}^q \left| \nabla_x f(\tilde{z}, \tilde{y}) + \lambda^T \nabla_y g(\tilde{z}, \tilde{y}) + \mu^T \nabla_y h(\tilde{z}, \tilde{y}) = \sigma_{\text{MFCQ}}, \lambda \nabla y g(\tilde{z}, \tilde{y}) = 0, \lambda \in \mathbb{R}^p \right. \right\}.
\]

Finally, we need to introduce a constraint qualification for optimal control problems of type (12). In order to define such a condition, we take a closer look at the control constraints in problem (1). Therefore, we define a mapping \( G: L_2^\infty[0, T] \rightarrow L_2^\infty[0, T] \) as follows:

\[
\forall u \in L_2^\infty[0, T], \forall t \in [0, T]: \quad G(u)(t) := G(t, u(t)).
\]
Furthermore, we need an ordering cone $K_0 \subseteq L_0^r[0, T]$ given by
\[ K_0 := \{ w \in L_0^r[0, T] \mid w(t) \geq 0 \text{ a.e. } t \in [0, T] \}. \]

Using these new terms we are able to equivalently express the constraint
\[ G(t, u(t)) \leq 0 \quad \text{a.e. } t \in [0, T] \]
of (1) by means of:
\[ \mathcal{G}(u) \in -K_0. \]

The following lemma from Bonnans (cf. [4]) presents some basic properties of the mapping $\mathcal{G}$ we need later in order to characterize certain multipliers appearing within the necessary optimality conditions.

**Lemma 4.3.** [4]
Consider an arbitrary point $\hat{u} \in L^m_0[0, T]$ and some $\lambda \in N_{-K_0}(\mathcal{G}(\hat{u}))$. Then the following statements hold true.

(i) $\mathcal{G}$ is continuously Fréchet differentiable at $\hat{u}$ and we have:
\[ \forall d_u \in L^m_0[0, T] \forall t \in [0, T]: \quad \mathcal{G}'(\hat{u})[d_u](t) = \nabla_u G(t, \hat{u}(t)) d_u(t). \]

(ii) If there exist a vector $a \in \mathbb{R}^r$ and a function $d_a \in L^m_0[0, T]$ such that
\[ G(t, \hat{u}(t)) + \nabla_u G(t, \hat{u}(t))d_a(t) + a \leq 0 \quad \text{a.e. } t \in [0, T] \]
\[ \mathcal{G}'(\hat{u})^*[a] \in L^m_0[0, T] \]
hold, then we have $\lambda \in L^r_0[0, T]$ (therein $\mathcal{G}'(\hat{u})^*$ denotes the adjoint operator of $\mathcal{G}'(\hat{u})$).

Now we are able to formulate a regularity condition for problem (1).

**Definition 4.1.** (Upper level regularity)
A feasible point $(\hat{x}, \hat{u}, \hat{y}) \in E_2$ of (1), (2) is called upper level regular if there exist $a \in \mathbb{R}^r$ and $\hat{u} \in L^m_0[0, T]$ such that the following conditions are satisfied (for almost every $t \in [0, T]$):

(i) $G(t, \hat{u}(t)) + \nabla_u G(t, \hat{u}(t)) (\hat{u}(t) - \hat{u}(t)) + a \leq 0$,

(ii) $\int_0^T \Phi^{-1}(t) \nabla_y F(t, \hat{x}(t), \hat{u}(t))(\hat{u}(t) - \hat{u}(t)) dt = 0$,

(iii) the linearized system
\[ \dot{x}(t) = \nabla_x F(t, \hat{x}(t), \hat{u}(t)) x(t) + \nabla_u F(t, \hat{x}(t), \hat{u}(t)) u(t) \quad (22) \]
is controllable,

(iv) $\nabla H(\hat{x}(T))$ possesses full row rank s.

Note that $\Phi$ in condition (ii) of the latter definition denotes the solution of the matrix differential equation (19). An obvious case where a feasible point $(\hat{x}, \hat{u}, \hat{y}) \in E_2$ of (1), (2) is upper level regular appears when no control constraints of the form $\mathcal{G}(u) \in -K_0$ appear, (22) is controllable and $\nabla H(\hat{x}(T))$ possesses full row rank s.

Next, we give a quite technical lemma which will be useful in order to prove the necessary optimality conditions for problem (1), (2). Therefore, we need to define a function $\omega_{\lambda, \hat{u}} \in W^1_{1, 0}[0, T]$ for some feasible point $(\hat{x}, \hat{u}, \hat{y}) \in E_2$ of (1), (2):
\[ \forall t \in [0, T]: \quad \omega_{\lambda, \hat{u}}(t) := \hat{x}(t) - \int_0^t \nabla_x F(\tau, \hat{x}(\tau), \hat{u}(\tau)) \hat{x}(\tau) + \nabla_u F(\tau, \hat{x}(\tau), \hat{u}(\tau)) \hat{u}(\tau) d\tau. \]
Lemma 4.4.
Let \( (\hat{x}, \hat{u}, \hat{y}) \in E_2 \) be a feasible point of (1), (2). Then the following statements hold.

(i) The function \( \hat{z} \in W_{0,\infty}^n[0, T] \) defined by

\[
\forall t \in [0, T]: \quad \hat{z}(t) := \hat{x}(t) - \Phi(t) \int_0^t \Phi^{-1}(\tau) \nabla_u F(\tau, \hat{x}(\tau), \hat{u}(\tau)) \hat{u}(\tau) d\tau
\]

is a solution of the Volterra equation (20) with \( \omega := \omega_{\hat{u}, \hat{y}} \).

(ii) Let \( (\hat{x}, \hat{u}, \hat{y}) \) be upper level regular, \( \hat{u} \) be the control satisfying the conditions (i) and (ii) of Definition 4.1 and \( \hat{z} \in W_{0,\infty}^n[0, T] \) be given as in statement (i) of this lemma. Furthermore, let \( \hat{x} \) be the solution of the linearized differential equation (22) with \( u := \hat{u} \) and initial condition \( x(0) = 0_{R^n} \). Then \( \hat{x} \) satisfies the terminal condition \( \hat{x}(T) = \hat{x}(T) - \hat{z}(T) \).

Proof:

(i) Fix some arbitrary \( t \in [0, T] \). Then we have from integration by parts and (19):

\[
\hat{z}(t) - \int_0^t \nabla_x F(\tau, \hat{x}(\tau), \hat{u}(\tau)) \hat{z}(\tau) d\tau
\]

\[
= \hat{x}(t) - \Phi(t) \int_0^t \Phi^{-1}(\tau) \nabla_u F(\tau, \hat{x}(\tau), \hat{u}(\tau)) \hat{u}(\tau) d\tau
\]

\[
- \int_0^t \nabla_x F(\tau, \hat{x}(\tau), \hat{u}(\tau)) \left[ \hat{x}(\tau) - \Phi(\tau) \int_0^{\tau} \Phi^{-1}(s) \nabla_u F(s, \hat{x}(s), \hat{u}(s)) \hat{u}(s) ds \right] d\tau
\]

\[
= \hat{x}(t) - \int_0^t \nabla_x F(\tau, \hat{x}(\tau), \hat{u}(\tau)) \hat{x}(\tau) d\tau - \Phi(t) \int_0^t \Phi^{-1}(\tau) \nabla_u F(\tau, \hat{x}(\tau), \hat{u}(\tau)) \hat{u}(\tau) d\tau
\]

\[
+ \int_0^t \nabla_x F(\tau, \hat{x}(\tau), \hat{u}(\tau)) \Phi(\tau) \left[ \int_0^\tau \Phi^{-1}(s) \nabla_u F(s, \hat{x}(s), \hat{u}(s)) \hat{u}(s) ds \right] d\tau
\]

\[
= \hat{x}(t) - \int_0^t \nabla_x F(\tau, \hat{x}(\tau), \hat{u}(\tau)) \hat{x}(\tau) d\tau - \Phi(t) \int_0^t \Phi^{-1}(\tau) \nabla_u F(\tau, \hat{x}(\tau), \hat{u}(\tau)) \hat{u}(\tau) d\tau
\]

\[
- \int_0^t \Phi(\tau) \left[ \int_0^{\tau} \Phi^{-1}(s) \nabla_u F(s, \hat{x}(s), \hat{u}(s)) \hat{u}(s) ds \right] d\tau
\]

\[
= \hat{x}(t) - \int_0^t \nabla_x F(\tau, \hat{x}(\tau), \hat{u}(\tau)) \hat{x}(\tau) d\tau - \Phi(t) \Phi^{-1}(\tau) \nabla_u F(\tau, \hat{x}(\tau), \hat{u}(\tau)) \hat{u}(\tau) d\tau
\]

\[
= \hat{x}(t) - \int_0^t \left[ \nabla_x F(\tau, \hat{x}(\tau), \hat{u}(\tau)) \hat{x}(\tau) + \nabla_u F(\tau, \hat{x}(\tau), \hat{u}(\tau)) \hat{u}(\tau) \right] d\tau = \omega_{\hat{u}, \hat{u}}(t).
\]

Hence, \( \hat{z} \) from (23) solves (20) with \( \omega := \omega_{\hat{u}, \hat{y}} \).
(ii) Observe that (since the boundary condition \( \tilde{x}(0) = \sigma_{2n} \) must hold) (22) possesses the unique solution \( \tilde{x} \) given as follows:

\[
\forall t \in [0, T]: \quad \tilde{x}(t) := \Phi(t) \int_0^t \Phi^{-1}(\tau) \nabla_u F(\tau, \tilde{x}(\tau), \tilde{u}(\tau)) d\tau.
\]

Therefore, we compute its terminal value using condition (ii) from the upper level regularity:

\[
\tilde{x}(T) = \Phi(T) \int_0^T \Phi^{-1}(t) \nabla_u F(t, \tilde{x}(t), \tilde{u}(t)) \tilde{u}(t) dt
\]

\[
= \Phi(T) \int_0^T \Phi^{-1}(t) \nabla_u F(t, \tilde{x}(t), \tilde{u}(t)) \tilde{u}(t) dt - \int_0^T \Phi^{-1}(t) \nabla_u F(t, \tilde{x}(t), \tilde{u}(t)) (\tilde{u}(t) - \tilde{u}(t)) dt
\]

\[
= \Phi(T) \int_0^T \Phi^{-1}(t) \nabla_u F(t, \tilde{x}(t), \tilde{u}(t)) \tilde{u}(t) dt = \hat{x}(T) - \tilde{x}(T).
\]

Hence, the proof is completed.

Now we can formulate the main result of this paper. Some ideas of the proof date back to Gerdts (cf. [11]) and Jahn (cf. [13]), who derived necessary optimality conditions for optimal control problems via the KKT-conditions in Banach spaces.

**Theorem 4.1. Linearized Pontryagin Maximum Principle**

Let \((\hat{x}, \hat{u}, \hat{y}) \in E_2\) be a local optimal solution of the bilevel programming problem (1), (2) such that MFCQ holds for problem (2) w.r.t. \( y \) in \((\hat{x}(T), \hat{y})\) and assume that \( \Psi \) is inner semicontinuous at this point. Furthermore, assume that \((\hat{x}, \hat{u}, \hat{y})\) is upper level regular. If (9) is partially calm at \((\hat{x}, \hat{u}, \hat{y})\), we can find functions \( \vartheta \in W^m_{1,\infty}[0, T] \) and \( \lambda \in K_0 \), a constant \( \rho > 0 \) and vectors \( \lambda_1, \lambda_2 \in \mathbb{R}^{p+}, \nu \in \mathbb{R}^q \) as well as \( \mu_1, \mu_2 \in \mathbb{R}^q \) such that the following conditions are satisfied (for almost every \( t \in [0, T] \)):

(i) Adjoint condition

\[
-\vartheta(t)^T = \vartheta(t)^T \nabla_x F(t, \hat{x}(t), \hat{u}(t)) - \nabla_x F(t, \hat{x}(t), \hat{u}(t), \hat{y}),
\]

(ii) Transversality condition

\[
\vartheta(T) = \nabla_x g(\hat{x}(T), \hat{y})^T (\rho \lambda_2 - \lambda_1) + \nabla_x h(\hat{x}(T), \hat{y})^T (\rho \mu_2 - \mu_1) - \nabla_x F(\hat{x}(T), \hat{y})^T - \nabla H(\hat{x}(T))^T \nu,
\]

(iii) Complementarity condition

\[
\lambda_1^T g(\hat{x}(T), \hat{y}) = 0 \quad \lambda(t)^T G(t, \hat{u}(t)) = 0,
\]

(iv) Lower level optimality condition

\[
\sigma_{2n} = \nabla_y f(\hat{x}(T), \hat{y})^T + \nabla_y g(\hat{x}(T), \hat{y})^T \lambda_2 + \nabla_y h(\hat{x}(T), \hat{y})^T \mu_2 - \lambda_2^T \lambda(\hat{x}(T), \hat{y}) = 0,
\]

(v) Linearized Pontryagin Maximum Principle

\[
\nabla_x F(t, \hat{x}(t), \hat{u}(t))^T \vartheta(t) - \nabla_x F(t, \hat{x}(t), \hat{u}(t), \hat{y})^T - \nabla_y G(t, \hat{u}(t))^T \lambda(t) = \sigma_{2n},
\]

(vi) Multiplier condition

\[
\nabla_y F(\hat{x}(T), \hat{y}) + \rho \cdot \nabla_y f(\hat{x}(T), \hat{y}) + \lambda_1^T \nabla_y g(\hat{x}(T), \hat{y}) + \mu_1^T \nabla_y h(\hat{x}(T), \hat{y}) = - \int_0^T \nabla_y F(\hat{x}(t), \hat{u}(t), \hat{y}) dt.
\]
Proof:

Observe that $(\hat{x}, \hat{u}, \hat{y})$ is a local optimal solution of problem (9) as well. Since this problem is partially calm at $(\hat{x}, \hat{u}, \hat{y})$, we make use of Lemma 3.1 to find $\rho > 0$ such that $(\hat{x}, \hat{u}, \hat{y})$ is a local optimal solution of problem (12). Now we are going to interpret problem (12) as a generalized Banach space optimization problem (3) so that we can use Theorem 2.1 in order to derive the necessary optimality conditions (i) to (vi).

Let us consider the spaces and sets $X := E_2$, $Y := \mathbb{R}^p \times L^\infty_0[0, T]$, $Z := W^q_{1,\infty}[0, T] \times \mathbb{R}^e \times \mathbb{R}^f$, $\Omega := E_2$ as well as $K := \mathbb{R}^p \times K_0$. Furthermore, we define functions $\psi: X \to \mathbb{R}$, $\alpha: X \to Y$ and $\beta: X \to Z$ as follows ($(x, u, y) \in X$ may be chosen arbitrarily):

$$\psi(x, u, y) := \rho(f(x(T), y) - \varphi(x(T))) + F_0(x(T), y) + \int_0^T F_1(t, x(t), u(t), y) dt$$

$$\alpha(x, u, y) := (g(x(T), y), G(u))$$

$$\beta(x, u, y) := \left\{x(\cdot) - x_0 - \int_0^T F(t, x(t), u(t)) dt, H(x(T)), h(x(T), y) \right\}.$$

Note that the condition $x(0) = x_0$ is used implicitly within the definition of $\beta$.

From Lemma 4.2 we have the local Lipschitz continuity of $\varphi$ around $\hat{x}(T)$ and we leave it to the reader to show that $\alpha$ and $\beta$ are continuously Fréchet differentiable (actually, this holds due to the choice of the infinity norm). We just give the corresponding subdifferentials and Fréchet derivatives here:

Let $(d_x, d_u, d_y) \in X$ be chosen arbitrarily. Then we have with Lemma 4.3:

$$d^T \psi(\hat{x}, \hat{u}, \hat{y}) | [d_x, d_u, d_y] = \left\{ \rho \cdot \nabla_x f(\hat{x}(T), \hat{y}) d_x(T) + \rho \cdot \nabla_y f(\hat{x}(T), \hat{y}) d_y(T) - \rho \cdot (d_x(T), \xi) \right.$$

$$\left. + \nabla F_0(\hat{x}(T), \hat{y}) d_x(T) + \nabla F_0(\hat{x}(T), \hat{y}) d_y(T) + \int_0^T \left[ \nabla_x F_1(t, \hat{x}(t), \hat{u}(t), \hat{y}) d_x(t) \right. \right.$$

$$\left. + \nabla_y F_1(t, \hat{x}(t), \hat{u}(t), \hat{y}) d_y(t) \right| \xi \in \partial \varphi(\hat{x}(T)) \right\}$$

$$d^T \alpha(\hat{x}, \hat{u}, \hat{y}) | [d_x, d_u, d_y] = \left\{ \nabla_x g(\hat{x}(T), \hat{y}) d_x(T) + \nabla_y g(\hat{x}(T), \hat{y}) d_y(T), \nabla_x \alpha(\hat{x}, \hat{u}) d_x(T) \right. \right.$$

$$\left. \left. + \nabla_y \alpha(\hat{x}, \hat{u}) d_y(T) \right| d_x(T) - \int_0^T \left[ \nabla_x F_1(t, \hat{x}(t), \hat{u}(t)) d_x(t) + \nabla_y F_1(t, \hat{x}(t), \hat{u}(t)) d_y(t) \right] dt, \right. \right.$$

$$\left. \nabla h(\hat{x}(T), \hat{y}) d_x(T), \nabla h(\hat{x}(T), \hat{y}) d_y(T) + \nabla \beta(\hat{x}, \hat{u}, \hat{y}) | [d_x, d_u, d_y] \right\}.$$

We show now that KRZCQ is satisfied at $(\hat{x}, \hat{u}, \hat{y})$. Since $K$ and $\Omega$ are sets with a nonempty interior, we just check MFCQ which is equivalent to KRZCQ due to Lemma 2.1.

We start by showing the surjectivity of $\beta'(\hat{x}, \hat{u}, \hat{y})$. Therefore, choose an arbitrary point $(v, v_u, v_y) \in Z$. Since $\nabla H(\hat{x}(T))$ possesses full row rank due to the upper level regularity we can find $w_x \in \mathbb{R}^s$ such that $\nabla H(\hat{x}(T)) w_x = v_x$ holds. Now consider the Volterra equation (20) where $\omega := v$ is satisfied. Choose one of its solutions and name it $v$. Take a look at the linearized dynamical system (22). Since it is controllable, we find a control $\hat{u}$ such that the corresponding solution $\hat{x}$ satisfies the boundary conditions $x(0) = x_0$ and $x(T) = w_x - u(T)$. We consider $\hat{x} := x + v$. Finally, we choose $\hat{y}$ as solution of the linear system $\nabla_y h(\hat{x}(T), \hat{y}) v_y = v_q - \nabla_y h(\hat{x}(T), \hat{y}) w_y$, which exists since $\nabla_y h(\hat{x}, \hat{y})$ has rank $q$ (this holds...
since MFCQ w.r.t. \( y \) is satisfied at \((\hat{x}(T), \hat{y})\) for problem (2)). Now observe the following:

\[
\beta'(\hat{x}, \hat{u}, \hat{y})[\bar{x}, \bar{u}, \bar{y}] = \left( \bar{x}(\cdot) - \int_0^T \nabla_x F(t, \hat{x}(t), \hat{u}(t))\hat{x}(t) + \nabla_u F(t, \hat{x}(t), \hat{u}(t))\hat{u}(t) \right) dt,
\]

\[
\nabla H(\hat{x}(T))\hat{x}(T), \nabla_x h(\hat{x}(T), \hat{y})\hat{x}(T) + \nabla_y h(\hat{x}(T), \hat{y})\hat{y}\right)
\]

\[
= \left( \bar{x}(\cdot) + \bar{u}(\cdot) - \int_0^T \nabla_x F(t, \hat{x}(t), \hat{u}(t))\hat{x}(t) + \nabla_u F(t, \hat{x}(t), \hat{u}(t))\hat{u}(t) \right) dt,
\]

\[
\nabla H(\hat{x}(T))\hat{x}(T), \nabla_x h(\hat{x}(T), \hat{y})\hat{x}(T) + \nabla_y h(\hat{x}(T), \hat{y})\hat{y}\right)
\]

\[
= \left( \bar{u}(\cdot) + \bar{y}(0), \nabla H(\hat{x}(T))\bar{w}_x + \bar{v}_y \right) = (\bar{u}, \bar{v}_y, \bar{v}_y).
\]

Since \((\bar{u}, \bar{v}_y, \bar{v}_y)\) was chosen arbitrarily, we deduce the surjectivity of \(\beta'(\hat{x}, \hat{u}, \hat{y})\).

Since MFCQ w.r.t. \( y \) holds at \((\hat{x}(T), \hat{y})\) for the lower level problem (2), we find some \( \bar{y} \in \mathbb{R}^k \) such that we have:

\[
g(\hat{x}(T), \hat{y}) + \nabla_x g(\hat{x}(T), \hat{y})(\hat{x}(T) - \bar{y}) < \sigma_{\mathbb{R}^k} \quad \nabla_y h(\hat{x}(T), \hat{y})(\hat{y} - \bar{y}) = 0_{\mathbb{R}^k}.
\]

(30)

Additionally, since \((\hat{x}, \hat{y}, \hat{u})\) is upper level regular, we find a control function \( \bar{u} \in L^\infty_{\mathbb{R}^k}[0, T] \) such that the solution \( \tilde{x} \) of (22) satisfies \( \tilde{x}(0) = \sigma_{\mathbb{R}^k} \) and \( \tilde{x}(T) = \bar{x}(T) - \hat{z}(T) \) where \( \hat{z} \) is defined in (23) (cf. statement (ii) of Lemma 4.4).

Furthermore, we have from the first condition of upper level regularity:

\[
\mathcal{G}(\bar{u}) + \mathcal{G}'(\bar{u})(\bar{u} - \tilde{u}) \in -\text{int}(K_0).
\]

Let us define \( \bar{x} := \bar{x} + \tilde{z} \). Then we (trivially) have \((\bar{x}, \bar{u}, \bar{y}) \in \text{int}(\Omega) \) and:

\[
\begin{align*}
\alpha(\bar{x}, \bar{u}, \bar{y}) &= \alpha'(\bar{x}, \bar{u}, \bar{y})[\bar{x}, \bar{u}, \bar{y}] = (\hat{x}(T) - \tilde{x}(T)) \cdot \mathcal{G}(\bar{u}) + \mathcal{G}'(\bar{u})(\bar{u} - \tilde{u}) \\
&= (\hat{x}(T) - \tilde{x}(T)) \cdot \mathcal{G}(\bar{u}) + \mathcal{G}'(\bar{u})(\bar{u} - \tilde{u}) \\
&= \mathcal{G}(\bar{u}) + \mathcal{G}'(\bar{u})(\bar{u} - \tilde{u}) \\
&\in (\mathbb{R}^k)^+ \times (\text{int}(K_0))^+ = -\text{int}(K).
\end{align*}
\]

Observe, that we have \(\beta'(\hat{x}, \hat{u}, \hat{y})[\bar{x}, \bar{u}, \bar{y}] = (\omega_{\tilde{u}}(\cdot), \nabla H(\hat{x}(T))\hat{x}(T), \nabla_x h(\hat{x}(T), \hat{y})\hat{x}(T) + \nabla_y h(\hat{x}(T), \hat{y})\hat{y})\hat{y}\) and hence

\[
\beta'(\hat{x}, \hat{u}, \hat{y})[\bar{x}, \bar{u}, \bar{y}] = \left( \bar{x}(\cdot) - \int_0^T \nabla_x F(t, \hat{x}(t), \hat{u}(t))\hat{x}(t) + \nabla_u F(t, \hat{x}(t), \hat{u}(t))\hat{u}(t) \right) dt,
\]

\[
\nabla H(\hat{x}(T))\hat{x}(T), \nabla_x h(\hat{x}(T), \hat{y})\hat{x}(T) + \nabla_y h(\hat{x}(T), \hat{y})\hat{y}\right)
\]

\[
= \left( \bar{x}(\cdot) + \bar{z}(\cdot) - \int_0^T \nabla_x F(t, \hat{x}(t), \hat{u}(t))\bar{z}(t) + \nabla_u F(t, \hat{x}(t), \hat{u}(t))\bar{u}(t) \right) dt
\]

\[
\nabla H(\hat{x}(T))\bar{z}(T), \nabla_x h(\hat{x}(T), \hat{y})\bar{z}(T) + \nabla_y h(\hat{x}(T), \hat{y})\bar{y}\right)
\]

17
\[ \omega_{\xi, \hat{v}}(\xi) + \hat{v}(0) - \int_0^T \left[ \nabla_x F(t, \hat{x}(t), \hat{u}(t)) \hat{x}(t) + \nabla_u F(t, \hat{x}(t), \hat{u}(t)) \hat{u}(t) \right] dt, \]

\[ \nabla H(\hat{x}(T))(\hat{x}(T) - \hat{z}(T) + \hat{z}(T)), \nabla h(\hat{x}(T), \hat{y})(\hat{x}(T) - \hat{z}(T) + \hat{z}(T)) + \nabla h(\hat{x}(T), \hat{y}) \hat{y} \]

which comes from the fact that \( \hat{z} \) solves the Volterra equation (20) with \( \omega := \omega_{\xi, \hat{v}} \). Summing up the conditions above we see that KRZCQ holds at \((\hat{x}, \hat{u}, \hat{y})\).

Now we use Theorem 2.1 (observe that its statements stay true even if \( \psi \) is locally Lipschitz continuous just in a neighbourhood of the reference point) to derive the existence of \( \xi \in \partial^p \psi(\hat{x}, \hat{u}, \hat{y}), (\lambda_1, \lambda) \in \mathbb{R}_{0}^{n^*} \times K_0^p \) and \((\mu, \nu, \mu_1) \in (W^{1,\infty}[0, T])^* \times \mathbb{R}^t \times \mathbb{R}^q \) such that

\[ \forall (x, u, y) \in E_2: \quad (\xi + \alpha'(\hat{x}, \hat{u}, \hat{y}) \circ (\lambda_1, \lambda) + \beta'(\hat{x}, \hat{u}, \hat{y}) \circ (\mu, \nu, \mu_1)) [x, u, y] = 0 \quad (31) \]

and

\[ 0 = \langle \alpha(\hat{x}, \hat{u}, \hat{y}), (\lambda_1, \lambda) \rangle \quad (32) \]

hold. Note that we we have equality in (31) since \( \Omega \) equals the whole space \( E_2 \). Additionally, (32) equals \( (G(\hat{u}), 0) = 0 \) and \( \lambda_1^p g(\hat{x}(T), \hat{y}) = 0 \). The latter condition is part of the complementarity condition (26). Since \((G(\hat{u}), 0) = 0 \) holds true, from the definition of the normal cone and the convexity of \( K_0 \) we deduce \( \lambda \in N_{\partial \psi_0}(G(\hat{u})) \).

From (31) and \( \xi \in \partial^p \psi(\hat{x}, \hat{u}, \hat{y}) \) we derive the existence of \( \xi \in \partial^p \psi(\hat{x}(T)) \) such that for any \((x, u, y) \in E_2\) we have:

\[ \rho \left( \nabla_x f(\hat{x}(T), \hat{y}) x(T) + \nabla_u f(\hat{x}(T), \hat{y}) y + \alpha'(\hat{x}, \hat{u}, \hat{y}) \circ (\lambda_1, \lambda) + \beta'(\hat{x}, \hat{u}, \hat{y}) \circ (\mu, \nu, \mu_1) \right) [x, u, y] = 0 \quad (33) \]

We make use of Lemma 4.2 to show the existence of \( \lambda_2 \in \mathbb{R}_{0}^{n^*} \) and \( \mu_2 \in \mathbb{R}^q \) such that (27) as well as

\[ \zeta = \nabla_x f(\hat{x}(T), \hat{y}) + \lambda_2^p \nabla g(\hat{x}(T), \hat{y}) + \mu_2^p \nabla h(\hat{x}(T), \hat{y}) \]

hold. That is why we have for any \((x, u, y) \in \Omega\):

\[ \rho \left( \nabla_x f(\hat{x}(T), \hat{y}) x(T) - \lambda_2^p \nabla g(\hat{x}(T), \hat{y}) x(T) - \mu_2^p \nabla h(\hat{x}(T), \hat{y}) y(T) \right) + \nabla_u F_0(\hat{x}(T), \hat{y}) x(T) + \nabla_y F_0(\hat{x}(T), \hat{y}) y(T) \]

\[ + \int_0^T \left[ \nabla_x F_1(t, \hat{x}(t), \hat{u}(t), \hat{y}) x(t) + \nabla_u F_1(t, \hat{x}(t), \hat{u}(t), \hat{y}) u(t) + \nabla_y F_1(t, \hat{x}(t), \hat{u}(t), \hat{y}) y(t) \right] dt \]

\[ + \lambda_1^p \left( \nabla_x g(\hat{x}(T), \hat{y}) x(T) + \nabla_y g(\hat{x}(T), \hat{y}) y(T) \right) + \alpha(\cdot)(u(\cdot), \lambda) \]

\[ + \int_0^T \left[ \nabla_x F_2(t, \hat{x}(t), \hat{u}(t), \hat{y}) x(t) + \nabla_u F_2(t, \hat{x}(t), \hat{u}(t), \hat{y}) u(t) + \nabla_y F_2(t, \hat{x}(t), \hat{u}(t), \hat{y}) y(t) \right] dt \]

\[ + \mu_1^p \left( \nabla_x h(\hat{x}(T), \hat{y}) x(T) + \nabla_y h(\hat{x}(T), \hat{y}) y(T) \right) = 0. \]
Taking $u = \sigma_{L^2[0,T]}$ and $y = \sigma_{R^2}$ we get from (33) for any $x \in W^m_{1,\infty}[0,T]$:  
\[
\left[\nabla_x F_0(\hat{x}(t), \hat{y}) + v^T \nabla H(\hat{x}(t)) + (\lambda_1 - \rho \lambda_2)^T \nabla_s g(\hat{x}(t), \hat{y}) + (\mu_1 - \rho \mu_2)^T \nabla_s h(\hat{x}(t), \hat{y})\right] x(t) \\
+ \int_0^T \nabla_x F_1(t, \hat{x}(t), \hat{u}(t), \hat{y}) x(t) dt + \left( x(\cdot) - \int_0^T \nabla_x F(t, \hat{x}(t), \hat{u}(t)) x(t) dt, \mu \right) = 0.
\]  
(34)

Furthermore, if we plug $x = \sigma_{W^m_{1,\infty}[0,T]}$ and $u = \sigma_{L^2[0,T]}$ in (33) we get for any $y \in R^k$:  
\[
\left[\nabla_y F_0(\hat{x}(t), \hat{y}) + \rho \cdot \nabla_y f(\hat{x}(t), \hat{y}) + \lambda_1^T \nabla_s g(\hat{x}(t), \hat{y}) + \mu_1^T \nabla_s h(\hat{x}(t), \hat{y}) + \int_0^T \nabla_y F_1(t, \hat{x}(t), \hat{u}(t), \hat{y}) dt\right] y = 0.
\]  
(35)

Obviously, the multiplier condition (29) is a consequence of (35).

Finally, we can choose $x = \sigma_{W^m_{1,\infty}[0,T]}$ as well as $y = \sigma_{R^2}$ in (33) in order to get  
\[
\int_0^T \nabla_u F_1(t, \hat{x}(t), \hat{u}(t), \hat{y}) u(t) dt + \left( \nabla_u G(\cdot, \hat{u}(\cdot)) u(\cdot), \lambda \right) + \left( - \int_0^T \nabla_u F(t, \hat{x}(t), \hat{u}(t)) u(t) dt, \mu \right) = 0
\]  
(36)

for any $u \in L^m_{\infty}[0,T]$.

We draw our attention to equation (34) in order to characterise the multiplier $\mu$. Since we want to write the following as short as possible, we define a vector $\xi \in R^n$ as given below:
\[
\xi^T := \nabla_x F_0(\hat{x}(t), \hat{y}) + v^T \nabla H(\hat{x}(t)) + (\lambda_1 - \rho \lambda_2)^T \nabla_s g(\hat{x}(t), \hat{y}) + (\mu_1 - \rho \mu_2)^T \nabla_s h(\hat{x}(t), \hat{y}).
\]

Now, we consider the Volterra equality (20) for an arbitrary $\omega \in W^m_{1,\infty}[0,T]$ and the (not necessarily unique) corresponding solution from (21). Using (34) we derive:
\[
\langle \omega, \mu \rangle = -\xi^T \left( \omega(T) + \Phi(T) \int_0^T \Phi^{-1}(t) \nabla_x F(t, \hat{x}(t), \hat{u}(t)) \omega(t) dt \right) - \int_0^T \left[ \nabla_x F_1(t, \hat{x}(t), \hat{u}(t), \hat{y}) \left( \omega(t) + \Phi(t) \int_0^t \Phi^{-1}(\tau) \nabla_x F(\tau, \hat{x}(\tau), \hat{u}(\tau)) \omega(\tau) d\tau \right) dt \right].
\]

We use integration by parts in order to derive for any $\omega \in W^m_{1,\infty}[0,T]$:  
\[
\langle \omega, \mu \rangle = -\xi^T \left( \omega(T) + \Phi(T) \int_0^T \Phi^{-1}(t) \nabla_x F(t, \hat{x}(t), \hat{u}(t)) \omega(t) dt \right) - \int_0^T \nabla_x F_1(t, \hat{x}(t), \hat{u}(t), \hat{y}) \omega(t) dt \\
\left. - \int_0^T \nabla_x F_1(t, \hat{x}(t), \hat{u}(t), \hat{y}) \Phi(t) dt \right|_0^T \\
+ \int_0^T \int_0^t \nabla_x F_1(\tau, \hat{x}(\tau), \hat{u}(\tau), \hat{y}) \Phi(\tau) d\tau \Phi^{-1}(t) \nabla_x F(t, \hat{x}(t), \hat{u}(t)) \omega(t) dt.
\]
We now define a function \( \varrho \in L^\infty_1[0, T] \) for any \( t \in [0, T] \) as follows:

\[
\varrho(t)^T := -\xi^T \Phi(T) \Phi^{-1}(t) \nabla_x F(t, \hat{x}(t), \hat{u}(t)) - \nabla_x F_1(t, \hat{x}(t), \hat{u}(t), \hat{y})
\]

\[
- \left( \int_0^T \nabla_x F_1(\tau, \hat{x}(\tau), \hat{u}(\tau), \hat{y}) \Phi(\tau) d\tau \right) \Phi^{-1}(t) \nabla_v F(t, \hat{x}(t), \hat{u}(t)) \omega(t) dt
\]

\[
= -\xi^T \omega(T) + \int_0^T \left[ -\xi^T \Phi(T) \Phi^{-1}(t) \nabla_x F(t, \hat{x}(t), \hat{u}(t)) - \nabla_x F_1(t, \hat{x}(t), \hat{u}(t), \hat{y})
\right]
\]

\[
- \left( \int_0^T \nabla_x F_1(\tau, \hat{x}(\tau), \hat{u}(\tau), \hat{y}) \Phi(\tau) d\tau \right) \Phi^{-1}(t) \nabla_v F(t, \hat{x}(t), \hat{u}(t)) \omega(t) dt
\]

Hence, we have:

\[
\forall \omega \in W^{n}_{1,\infty}[0, T]: \quad \langle \omega, \mu \rangle = -\xi^T \omega(T) + \int_0^T \varrho(t)^T \omega(t) dt.
\] (37)

Let \( t \in [0, T] \) be chosen arbitrarily. Then we have \( I_n = \Phi(t) \Phi^{-1}(t) \) and from the product rule we deduce:

\[
\Omega^{n,0} = \nabla [\Phi(\cdot) \Phi^{-1}(\cdot)]^T(\cdot) = \Phi(t) \Phi^{-1}(t) + \Phi(t)(\Phi^{-1}(t))'.
\]

A look at the definition of \( \Phi \) in (19) now reveals for almost every \( t \in [0, T] \):

\[
(\Phi^{-1}(t)) = -\Phi^{-1}(t)(\Phi(t))^{-1}(t) = -\Phi^{-1}(t) \nabla_x F(t, \hat{x}(t), \hat{u}(t)) \Phi(t) \Phi^{-1}(t) = -\Phi^{-1}(t) \nabla_x F(t, \hat{x}(t), \hat{u}(t)).
\]

Plugging this into the definition of \( \varrho \) shows for almost every \( t \in [0, T] \):

\[
\varrho(t)^T = \xi^T \Phi(T)(\Phi^{-1}(t)) - \nabla_x F_1(t, \hat{x}(t), \hat{u}(t), \hat{y}) + \left( \int_0^T \nabla_x F_1(\tau, \hat{x}(\tau), \hat{u}(\tau), \hat{y}) \Phi(\tau) d\tau \right) \Phi^{-1}(t).
\]

Now we define the function \( \vartheta \in W^{n}_{1,\infty}[0, T] \) for any \( t \in [0, T] \) as follows:

\[
\vartheta(t)^T := -\xi^T \Phi(T) \Phi^{-1}(t) - \left( \int_0^T \nabla_x F_1(\tau, \hat{x}(\tau), \hat{u}(\tau), \hat{y}) \Phi(\tau) d\tau \right) \Phi^{-1}(t).
\]

Observe that from the product rule of differentiation we have:

\[
\dot{\vartheta}(t)^T = -\xi^T \Phi(T)(\Phi^{-1}(t)) - \left[ -\nabla_x F_1(t, \hat{x}(t), \hat{u}(t), \hat{y}) \Phi(t) \right] \Phi^{-1}(t)
\]

\[
- \left( \int_0^T \nabla_x F_1(\tau, \hat{x}(\tau), \hat{u}(\tau), \hat{y}) \Phi(\tau) d\tau \right) \Phi^{-1}(t) = -\varrho(t)^T.
\]

20
From the definition of $\vartheta$ we have $\vartheta(T) = -\xi^T$. Recalling the definition of $\xi$ we see that this equals the transversality condition (25). Furthermore, we have

$$
\vartheta(t)^T \nabla_x F(t, \dot{x}(t), \dot{u}(t)) - \nabla_x F(t, \dot{x}(t), \dot{u}(t), \dot{y}) = -\xi^T \Phi(T)\Phi^{-1}(t)\nabla_x F(t, \dot{x}(t), \dot{u}(t), \dot{y})
$$

and hence the adjoint condition (24) holds.

From (37) we deduce:

$$
\left| \int \nabla_x F(t, \dot{x}(t), \dot{u}(t)) u(t) dt \right| = \left| \int \nabla_x F(t, \dot{x}(t), \dot{u}(t), \dot{y}) u(t) dt \right|
$$

and hence the adjoint condition (24) holds.

From (37) we deduce:

$$
\forall \omega \in W^1_{1,\infty}[0, T]: \quad \langle \omega, \mu \rangle = \vartheta(T)^T \omega(T) - \int_0^T \vartheta(t)^T \omega(t) dt.
$$

(38)

Now we are able to characterise the multiplier $\lambda$ which appears in (36). We rearrange the latter equation and use integration by parts again to derive from (38) for any $u \in L^\infty_m[0, T]$:

$$
\left\langle \nabla_u G(\cdot, \dot{u}(\cdot))u(\cdot), \lambda \right\rangle = \left( \int_0^T \nabla_u F(t, \dot{x}(t), \dot{u}(t)) u(t) dt \right) - \int_0^T \nabla_u F(t, \dot{x}(t), \dot{u}(t), \dot{y}) u(t) dt
$$

$$
= \vartheta(T)^T \left( \int_0^T \nabla_u F(t, \dot{x}(t), \dot{u}(t)) u(t) dt \right) - \int_0^T \vartheta(t)^T \left( \int_0^t \nabla_u F(t, \dot{x}(t), \dot{u}(t)) u(\tau) d\tau \right) dt
$$

$$
- \int_0^T \nabla_u F_1(t, \dot{x}(t), \dot{u}(t), \dot{y}) u(t) dt
$$

$$
= \vartheta(T)^T \left( \int_0^T \nabla_u F(t, \dot{x}(t), \dot{u}(t)) u(t) dt \right) - \int_0^T \nabla_u F_1(t, \dot{x}(t), \dot{u}(t), \dot{y}) u(t) dt
$$

$$
- \vartheta(t)^T \left( \int_0^t \nabla_u F(t, \dot{x}(t), \dot{u}(t)) u(\tau) d\tau \right)dt + \vartheta(t)^T \nabla_u F_1(t, \dot{x}(t), \dot{u}(t), \dot{y}) u(t) dt
$$

$$
= \int_0^T \vartheta(t)^T \nabla_u F(t, \dot{x}(t), \dot{u}(t)) - \nabla_u F_1(t, \dot{x}(t), \dot{u}(t), \dot{y}) u(t) dt.
$$

Observe that the mapping $t \mapsto \nabla_x F(t, \dot{x}(t), \dot{u}(t))^T \vartheta(t) - \nabla_x F_1(t, \dot{x}(t), \dot{u}(t), \dot{y})^T$ belongs to the space $L^\infty_m[0, T]$. Hence, let $C \geq 0$ denote its norm in this space. As a consequence we have

$$
\left| \left\langle u, G(\cdot, \dot{u}(\cdot))u(\cdot), \lambda \right\rangle \right| = \int_0^T \left| \vartheta(t)^T \nabla_u F(t, \dot{x}(t), \dot{u}(t)) - \nabla_u F_1(t, \dot{x}(t), \dot{u}(t), \dot{y}) \right| u(t) dt
$$

$$
\leq C \cdot \sum_{i=1}^m \int_0^T |e_i^T u(t)| dt = C \cdot \sum_{i=1}^m \|e_i^T u\|_{L^1[0, T]} \leq C \cdot m \cdot \|u\|_{L^1[0, T]}
$$

21
where \( e_i \in \mathbb{R}^m \) denotes the \( i \)-th unity vector in \( \mathbb{R}^m \). That is why the linear operator \( u \mapsto \langle u, G'(\hat{u}) \rangle [\lambda] \) is bounded (and hence continuous) w.r.t. the \( L_1 \)-norm and consequently it can be interpreted as an element of the dual space \((L^m_\infty[0, T])^*\) which equals \( L^m_\infty[0, T] \). From statement (ii) of Lemma 4.3 and the first condition of upper level regularity we now deduce \( \lambda \in L^m_\infty[0, T] \). That is why we can write

\[
\langle \nabla_u G(\cdot, \hat{u}(\cdot)) u(\cdot), \lambda \rangle = \int_0^T \lambda(t)^T \nabla_u G(t, \hat{u}(t)) u(t) dt
\]

and especially

\[
\int_0^T \left[ \theta(t)^T \nabla_x F(t, \hat{x}(t), \hat{u}(t)) - \nabla_x F_1(t, \hat{x}(t), \hat{u}(t), \hat{y}) - \lambda(t)^T \nabla_x G(t, \hat{u}(t)) \right] u(t) dt = 0
\]

for all \( u \in L^m_\infty[0, T] \). The latter equation implies the Linearized Pontryagin Maximum Principle stated in (28). Furthermore, \( (G(\hat{u}), \lambda) = 0 \) from (32) can now be expressed equivalently by \( \lambda(t)^T G(t, \hat{u}(t)) = 0 \) for almost every \( t \in [0, T] \) which equals the second condition in (26). Additionally, from \( \lambda \in \mathcal{N}_{K_0}(G(\hat{u})) \cap L^m_\infty[0, T] = (K_0 + [G(\hat{u})])^D \cap L^m_\infty[0, T] \) we have \( \lambda \in K_0 \). Hence, the proof is completed. 

We can specify the obtained optimality conditions in terms of a linear lower level problem (2) and a linear dynamical system in (1) with vanishing control and linear endpoint constraints.

**Theorem 4.2.**
Consider the bilevel programming problem (1), (2) and assume

(i) there exist matrices \( A \in \mathbb{R}^{p \times n}, B \in \mathbb{R}^{p \times m}, C \in \mathbb{R}^{q \times n} \) and \( d \in \mathbb{R}^q \) such that we have for all \( t \in \mathbb{R}, x \in W^m_{1,n_0}[0, T] \) and \( u \in L^m_\infty[0, T] \):

\[
F(t, x(t), u(t)) := Ax(t) + Bu(t) \quad H(x(T)) := Cx(T) - d,
\]

(ii) the function \( G \) vanishes,

(iii) there exist matrices \( P \in \mathbb{R}^{p \times p}, Q \in \mathbb{R}^{q \times q}, R \in \mathbb{R}^{q \times k}, S \in \mathbb{R}^{k \times k}, a \in \mathbb{R}^p, b \in \mathbb{R}^q \) and \( c \in \mathbb{R}^k \) such that for all \( z \in \mathbb{R}^p \) and \( y \in \mathbb{R}^q \) the functions \( f, g \) and \( h \) possess the following structure:

\[
f(z, y) := c^T y \quad g(z, y) := P z + Q y - a \quad h(z, y) := R z + S y - b.
\]

Furthermore, assume that the controllability matrix

\[
\mathcal{J} := [B | AB | \ldots | A^{n-1}B] \in \mathbb{R}^{n \times m}
\]

has rank \( n \) and that \( C \) possesses rank \( s \) while \( S \) possesses rank \( q \).

If \((\hat{x}, \hat{u}, \hat{y}) \in E_2\) is a local optimal solution of (1), (2) such that we find \( \tilde{y} \in \mathbb{R}^k \) satisfying \( P \tilde{y}(T) + Q \tilde{y} < a \) and

\[
R \tilde{y}(T) + S \tilde{y} = b,
\]

then there exist a function \( \theta \in W^m_{1,n_0}[0, T] \), a constant \( \rho > 0 \) and vectors \( \lambda_1, \lambda_2 \in \mathbb{R}^{p \times \nu}, \nu \in \mathbb{R}^q \) as well as \( \mu_1, \mu_2 \in \mathbb{R}^q \) such that the following conditions are satisfied (for almost every \( t \in [0, T] \)):

(i) Adjoint condition

\[
-\dot{\theta}(t)^T = \theta(t)^T A - \nabla_x F_1(t, \hat{x}(t), \hat{u}(t), \hat{y}),
\]

(ii) Transversality condition

\[
\theta(T) = P^T (\rho \lambda_2 - \lambda_1) + R^T (\rho \mu_2 - \mu_1) - \nabla_x F_0(\hat{x}(T), \hat{y})^T - C^T \nu,
\]

(iii) Complementarity condition

\[
\lambda_1^T (P \tilde{y}(T) + Q \tilde{y} - a) = 0.
\]
Example 4.1. We now give a small example to illustrate Theorem 4.2. Apply the necessary optimality conditions (24) to (29) to the special case given in this theorem to obtain the conditions

\[ \text{with the notations of Theorem 4.2 we have} \]

\[ \text{and is controllable since the controllability matrix} \]

\[ \text{and the existence of} \ ̃y \text{ satisfying } P ̃x(T) + ̃Q ̃y < a \text{ and } R ̃x(T) + S ̃y = b \text{ we deduce that MFCQ w.r.t. } y \text{ holds for problem (2) at } (̃x(T), ̃y). \]

From [1] and [16] we conclude that \( Ψ \) is inner semicontinuous at \(( ̃x(T), ̃y)\) since problem (2) is fully linear. Furthermore, the system (22) takes the form

\[ \dot{x}(t) = Ax(t) + Bu(t) \]

and is controllable since the controllability matrix \( J \) presented in (39) is of rank \( n \) (cf. Theorem 4.1 in [2]). Thus, observing that \( G \) vanishes and \( C \) possesses rank \( s \), the point \(( ̇x, ̇y)\) is upper level regular (condition (ii) follows easily choosing \( ̃u \). Since the lower level problem (2) is of type (17), it possesses a uniformly weak sharp minimum (Lemma 3.4). That is why (9) is partially calm at \(( ̇x, ̇y)\) due to Remark 3.1 (observe that MFCQ w.r.t. \( y \) for problem (2) guarantees \( ̃x(T) ∈ \text{int(dom(Ψ))} \)). Consequently, all assumptions of Theorem 4.1 hold and we can apply the necessary optimality conditions (24) to (29) to the special case given in this theorem to obtain the conditions (40) to (45).

Proof:

From the rank condition on \( S \) and the existence of \( ̃y \) satisfying \( P ̃x(T) + ̃Q ̃y < a \) and \( R ̃x(T) + S ̃y = b \) we deduce that MFCQ w.r.t. \( y \) holds for problem (2) at \(( ̃x(T), ̃y)\). From [1] and [16] we conclude that \( Ψ \) is inner semicontinuous at \(( ̃x(T), ̃y)\) since problem (2) is fully linear. Furthermore, the system (22) takes the form

\[ \dot{x}(t) = Ax(t) + Bu(t) \]

We now give a small example to illustrate Theorem 4.2.

Example 4.1. Let us consider the following optimistic bilevel optimal control problem:

\[ \int_0^1 (u(t) - y)^2 dt \to \min_{u,y} \]

\[ \dot{x}(t) = u(t) \quad \text{a.e. } t ∈ [0, 1] \]

\[ x(0) = 0 \]

\[ x(1) = 1 \]

\[ y ∈ \text{Argmin}_{y} \{ y | x(1) - y ≤ 0 \}. \]

Simple calculations show that a global optimal solution of (46) is \(( ̂x, ̂u, ̂y)\) such that \( ̂x(t) = t, ̂u(t) = 1 \) and \( ̂y = 1 \) holds for all \( t ∈ [0, 1] \). We show that the assumptions of Theorem 4.2 are satisfied and depict that the necessary optimality conditions hold at \(( ̂x, ̂u, ̂y)\).

With the notations of Theorem 4.2 we have \( A = 0, B = 1, C = 1, d = 1, P = 1, Q = -1, R = 0, S = 0, a = 0, b = 0, c = 1, F_0(x(T), y) ≡ 0 \) and \( F_1(t, x(t), u(t), y) = (u(t) - y)^2 \) for \(( t, x, u, y) ∈ IR × W_{1,∞}[0, 1] × L_{∞}[0, 1] × IR \). Note that dom(Ψ) = IR holds since \( Ψ(x(1)) = \{ x(1) \} \) is satisfied for any \( x ∈ W_{1,∞}[0, 1] \). Additionally, the controllability matrix \( J \) from (39) takes the form \( J = 1 \) and has full row rank 1. Furthermore, the matrix \( C \) possesses full row rank 1. Moreover, we can choose \( ̃y = 2 \) to see that MFCQ w.r.t. \( y \) holds for problem (2) at \(( ̃x(1), ̃y)\). That is why all the assumptions of Theorem 4.2 hold. That means we can find \( θ ∈ W_{1,∞}[0, 1] \) numbers \( λ_1, λ_2 ≥ 0, ν ∈ IR \) and \( ρ > 0 \) such that the following conditions are satisfied (for almost every \( t ∈ [0, 1] \)):
(i) Adjoint condition
\(-\dot{v}(t) = 0,\)

(ii) Transversality condition
\(\dot{v}(1) = 1 \cdot (\rho \lambda_2 - \lambda_1) - 1 \cdot \nu;\)

(iii) Complementarity condition
\(\lambda_1 (1 \cdot 1 + (-1) \cdot 1 - 0) = 0,\)

(iv) Lower level optimality condition
\(1 + (-1) \lambda_2 = 0 \quad \lambda_2 (1 \cdot 1 + (-1) \cdot 1 - 0) = 0,\)

(v) Linearized Pontryagin Maximum Principle
\(1 \cdot \dot{v}(t) - 2(1 - 1) = 0;\)

(vi) Multiplier condition
\(\rho \cdot 1 + \lambda_1 \cdot (-1) = -\int_0^1 (-2)(1 - 1) d\tau = 0.\)

From (iv) we deduce \(\lambda_2 = 1,\) (vi) reveals \(\rho = \lambda_1\) and hence we have \(\dot{v}(1) = -\nu\) from (ii). Choosing \(\nu = 0\) does not contradict \(\dot{v} \equiv 0\) which follows obviously from (v). Note that this choice of \(\dot{v}\) satisfies the adjoint condition (i) as well. Moreover, one may fix \(\rho = \lambda_1 > 0\). Since the complementarity condition (iii) holds automatically, the necessary optimality conditions of Theorem 4.2 hold.

We now want to give another constraint qualification for the upper level problem which implies the upper level regularity condition introduced in Definition 4.1 since the search for a control function \(\tilde{u}\) satisfying conditions (i) and (ii) of the latter definition could be time-consuming (the conditions (iii) and (iv) are standard in optimal control). Therefore, we go back to the proof of Theorem 4.1 and try to find a criterion which ensures that KRZCQ holds for the generalized Banach space optimization problem we used there. Take a look at the following definition.

**Definition 4.2. (Strong upper level regularity)**

Let \((\hat{x}, \hat{u}, \hat{\nu}) \in E_2\) be a feasible point of problem (1), (2). It is called strongly upper level regular if we can guarantee that the following conditions hold:

(i) \(\nabla_u G(t, \hat{u}(t))\) has rank \(r\) for every \(t \in [0, T]\),

(ii) there exists \(\epsilon \in \mathbb{R}^{r+}\) satisfying \(\nabla_u F(t, \hat{x}(t), \hat{u}(t))\nabla_u G(t, \hat{u}(t)) = \nabla_u G(t, \hat{u}(t))\nabla_u F(t, \hat{x}(t), \hat{u}(t))^T \nabla_u G(t, \hat{u}(t))^T)^{-1}\) \(= 0\) for almost every \(t \in [0, T]\),

(iii) the linearized system (22) is controllable,

(iv) the matrix \(\nabla H(\hat{x}(T))\) possesses full row rank \(s\).

It is obvious that condition (ii) of the above definition holds automatically, if for almost every \(t \in [0, T]\) we have

\[\nabla_u F(t, \hat{x}(t), \hat{u}(t))\nabla_u G(t, \hat{u}(t))^T = O^{n \times r}.\]

Note that this is just a sufficient condition for (ii). One may think of a feasible point \((\hat{x}, \hat{u}, \hat{\nu}) \in E_2\) of (1), (2) satisfying

\[
\nabla_u F(t, \hat{x}(t), \hat{u}(t)) = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \quad \nabla_u F(t, \hat{x}(t), \hat{u}(t)) = \begin{pmatrix} 2 & -6 \\ -2 & 6 \end{pmatrix} \quad \nabla_u G(t, \hat{u}(t)) = \begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix}
\]

for all \(t \in [0, T]\) \((H\text{ may vanish})\). This point is strongly upper level regular (choose \(\epsilon = \begin{pmatrix} 1 & 1 \end{pmatrix}^T\)) but for all \(t \in [0, T]\) we have \(\nabla_u F(t, \hat{x}(t), \hat{u}(t))\nabla_u G(t, \hat{u}(t))^T \neq O^{2 \times 2}.\)
Recall that the matrix $\nabla_x G(t, \hat{u}(t))^T \left[ \nabla_x G(t, \hat{u}(t)) \nabla_x G(t, \hat{u}(t))^T \right]^{-1}$ is the Moore-Penrose inverse of $\nabla_x G(t, \hat{u}(t))$ provided that the latter matrix possesses full row rank $r$ (which is satisfied if $(\hat{x}, \hat{u}, \hat{y})$ is a strongly upper level regular point of (1), (2)). Indeed, it follows from the definition that any strongly upper level regular point $(\hat{x}, \hat{u}, \hat{y}) \in E_2$ of (1), (2) is upper level regular.

**Lemma 4.5.**
If a feasible point $(\hat{x}, \hat{u}, \hat{y}) \in E_2$ is strongly upper level regular for problem (1), (2), then $(\hat{x}, \hat{u}, \hat{y})$ is upper level regular.

**Proof:**
Take $\epsilon \in \mathbb{R}^{m_p}$ from the definition of strong upper level regularity and define a function $\tilde{v} \in L^\infty_{\text{loc}}[0, T]$ as follows:

$$\forall t \in [0, T]: \quad \tilde{v}(t) = -\nabla_x G(t, \hat{u}(t))^T \left[ \nabla_x G(t, \hat{u}(t)) \nabla_x G(t, \hat{u}(t))^T \right]^{-1} \epsilon.$$ 

Note that $\nabla_x G(t, \hat{u}(t)) \nabla_x G(t, \hat{u}(t))^T$ is a regular matrix for every $t \in [0, T]$ since we assumed that $\nabla_x G(t, \hat{u}(t))$ has full row rank $r$. Observe that we have for almost every $t \in [0, T]$:

$$G(t, \hat{u}(t)) + \nabla_x G(t, \hat{u}(t)) \tilde{v}(t) = G(t, \hat{u}(t)) - \nabla_x G(t, \hat{u}(t)) \nabla_x G(t, \hat{u}(t))^T \left[ \nabla_x G(t, \hat{u}(t)) \nabla_x G(t, \hat{u}(t))^T \right]^{-1} \epsilon = G(t, \hat{u}(t)) - \epsilon.$$ 

Hence we have:

$$G(t, \hat{u}(t)) + \nabla_x G(t, \hat{u}(t)) \tilde{v}(t) + \epsilon = G(t, \hat{u}(t)) \leq 0_{\mathbb{R}}$$ 

a.e. $t \in [0, T]$. (47)

That means defining $\tilde{\hat{u}} := \hat{u} + \tilde{v}$ produces a control function that satisfies the first condition of upper level regularity with $a := \epsilon$. From the second condition of strong upper level regularity we have for almost every $t \in [0, T]$:

$$\nabla_x F(t, \hat{x}(t), \hat{u}(t))(\hat{u}(t) - \tilde{\hat{u}}(t)) = \nabla_x F(t, \hat{x}(t), \hat{u}(t)) \tilde{v}(t) = -\nabla_x F(t, \hat{x}(t), \hat{u}(t)) \nabla_x G(t, \hat{u}(t)) \tilde{v}(t) = -\nabla_x F(t, \hat{x}(t), \hat{u}(t)) \nabla_x G(t, \hat{u}(t))^T \left[ \nabla_x G(t, \hat{u}(t)) \nabla_x G(t, \hat{u}(t))^T \right]^{-1} \epsilon = 0_{\mathbb{R}}.$$ 

Hence, the integral condition (ii) of upper level regularity holds. The third and forth condition of Definition 4.1 equal the third and forth one of strong upper level regularity. Consequently, $(\hat{x}, \hat{u}, \hat{y})$ is upper level regular. #

It follows easily from Lemma 4.5 that the statements of Theorem 4.1 stay true under strong upper level regularity.

**Remark 4.1.**
Let $(\hat{x}, \hat{u}, \hat{y}) \in E_2$ be a local optimal solution of the bilevel programming problem (1), (2). Furthermore, let $(\hat{x}, \hat{u}, \hat{y})$ be a strongly upper level regular point such that MFCQ w.r.t. $y$ holds at $(\hat{x}(T), \hat{y})$ for problem (2). Moreover, let $\Psi$ be inner semicontinuous at $(\hat{x}(T), \hat{y})$. If (9) is partially calm at $(\hat{x}, \hat{u}, \hat{y})$, we can find functions $\vartheta \in W^1_{\text{loc}}[0, T]$ as well as $\lambda \in K_0$, a constant $\rho > 0$ and vectors $\lambda_1, \lambda_2 \in \mathbb{R}^{p_+}$, $v \in \mathbb{R}^s$ as well as $\mu_1, \mu_2 \in \mathbb{R}^q$ so that the conditions (i) to (vi) of Theorem 4.1 are satisfied.

Finally, we want to consider another small example to visualize the strong upper level regularity.
Example 4.2.
Consider the following optimistic bilevel optimal control problem:

\[
\int_0^1 \left[ (u_1(t) + y_1(t))^2 + 2(2u_2(t) - y_2(t))^2 \right] dt \to \min_{x, u, y}
\]

\[
\begin{pmatrix}
\dot{x}_1(t) \\
\dot{x}_2(t)
\end{pmatrix} = \begin{pmatrix}
0 & 0 \\
1 & 0
\end{pmatrix}\begin{pmatrix}
x_1(t) \\
x_2(t)
\end{pmatrix} + \begin{pmatrix}
1 & 1 \\
1 & 1
\end{pmatrix}\begin{pmatrix}
u(t) \\
u(t)
\end{pmatrix}
\quad a.e. \ t \in [0, 1]
\]

\[
x_1(0) = 1 \\
x_2(0) = 1
\]

\[
u_1(t) + \nu_2(t) \leq 2 \quad a.e. \ t \in [0, 1]
\]

\[
\begin{pmatrix}
x_1(1) \\
x_2(1)
\end{pmatrix}^2 + (x_2(1))^2 = 5
\]

\[
y \in \text{Argmin}_y \left\{ y_1 + y_2 \left| \begin{array}{c}
x_1(1) - y_1 \\
x_2(1) - y_2
\end{array} \leq 0 \right. \right\}.
\]

It is easy to see that a global optimal solution of (48) is given by \( \hat{x}_1(t) = 1, \hat{x}_2(t) = \hat{t} + 1, \hat{u}_1(t) = -1, \hat{u}_2(t) = 1, y_1 = 1 \) and \( y_2 = 2 \) for arbitrary \( t \in [0, 1] \). Since the point \( (\hat{x}, \hat{u}, \hat{y}) \) satisfies all the assumptions of Remark 4.1, the necessary optimality conditions hold at this point.

We start verifying that \( (\hat{x}, \hat{u}, \hat{y}) \) is strongly upper level regular. Obviously, we have \( \nabla_u G(t, \hat{u}(t)) = (-1, 1) \) and this matrix has full row rank \( r = 1 \). Additionally, we compute:

\[
\nabla_u F(t, \hat{x}(t), \hat{u}(t))\nabla_u G(t, \hat{u}(t))^T = \begin{pmatrix}
1 & 1 \\
1 & 1
\end{pmatrix}\begin{pmatrix}
-1 \\
1
\end{pmatrix} = \begin{pmatrix}
0 \\
0
\end{pmatrix}.
\]

The latter condition is sufficient for the second condition of strong upper level regularity. Moreover, the dynamical system of (48) is linear while its controllability matrix takes the form

\[
\mathcal{F} = \begin{pmatrix}
1 & 1 & 0 & 0 \\
1 & 1 & 1 & 1
\end{pmatrix}
\]

and has full row rank \( n = 2 \). Consequently, (22) is controllable. Finally, observe that we have \( \nabla H(\hat{x}(T)) = (2, 4) \) and this matrix possesses full row rank \( s = 1 \). That is why the point \( (\hat{x}, \hat{u}, \hat{y}) \) is strongly upper level regular.

Due to the very easy structure of the lower level problem we easily obtain \( \Psi(x(T)) = [x(T)] \) for every \( x \in W_{1, \infty}^2[0, 1] \) so that \( \Psi \) is continuous and hence inner semicontinuous at any point of its graph. Furthermore, the lower level problem has the structure of problem (17) which implies that the optimal value reformulation (9) of (48) is partially calm at \( (\hat{x}, \hat{u}, \hat{y}) \) (cf. Remark 3.1). Finally, we have \( \nabla_y g(\hat{x}(T), \hat{y}) = -I_2 \) and hence MFCQ holds w.r.t. \( y \) at \( (\hat{x}(T), \hat{y}) \) for the lower level problem.

From Remark 4.1 we get the existence of functions \( \theta \in W_{1, \infty}^2[0, 1] \) as well as \( \lambda \in L_{\infty}[0, 1] \), a constant \( \rho > 0 \), vectors \( \lambda_1, \lambda_2 \in \mathbb{R}_{\geq 0}^2 \) and \( \nu \in \mathbb{R} \) such that (24) to (29) hold. Namely, these multipliers must solve the following system (for almost every \( t \in [0, 1] \)):

(i) Adjoint condition

\[
-\hat{\theta}_1(t) = \theta_2(t) \quad -\hat{\theta}_2(t) = 0,
\]

(ii) Transversality condition

\[
\theta_1(1) = 1 \cdot (\rho \lambda_1^1 - \lambda_1^1) - 2 \nu \quad \theta_2(1) = 1 \cdot (\rho \lambda_2^2 - \lambda_2^1) - 4 \nu,
\]

(iii) Complementarity condition

\[
\lambda_1^1 (1 - 1) + \lambda_2^2 (2 - 2) = 0 \quad \lambda(t)(-1 + 1 - 2) = 0,
\]

\[
\lambda(t)(-(-1) + 1 - 2) = 0,
\]
(iv) Lower level optimality condition

\[ 0 = 1 + \lambda_2^1(-1) \quad 0 = 1 + \lambda_2^2(-1) \quad \lambda_2^1(1 - 1) + \lambda_2^2(2 - 2) = 0, \]

(v) Linearized Pontryagin Maximum Principle

\[ \vartheta_1(t) + \vartheta_2(t) - 2((-1) + 1) - (-1) \cdot \lambda(t) = 0 \quad \vartheta_1(t) + \vartheta_2(t) - 4(2 \cdot 1 - 2) - 1 \cdot \lambda(t) = 0, \]

(vi) Multiplier condition

\[ \rho \cdot 1 + \lambda_1^1(-1) = -\frac{1}{0} 2((-1) + 1)dt = 0 \quad \rho \cdot 1 + \lambda_1^2(-1) = -\frac{1}{0} (-2) \cdot (2 \cdot 1 - 2)dt = 0. \]

From (iv) we have \( \lambda_2^1 = \lambda_2^2 = 1 \) and using (vi) we deduce \( \lambda_1^1 = \lambda_1^2 = \rho \). Choosing \( \nu = 0 \) these facts cause \( \vartheta_1(1) = \vartheta_2(1) = 0 \) since the transversality condition (ii) must hold. Obviously, \( \vartheta_2 \) has to be constant (from the adjoint condition (i)) so that we have \( \vartheta_2 \equiv 0 \). Now from the Linearized Pontryagin Maximum Principle (v) we deduce \( \vartheta_1(t) + \vartheta_2(t) = 0 \) and \( \lambda(t) = 0 \) for almost every \( t \in [0, 1] \). Hence, we conclude \( \vartheta_1 \equiv 0 \) and \( \lambda \equiv 0 \). This does not contradict the remaining adjoint condition in (i). The choice \( \rho = \lambda_1^1 = \lambda_1^2 = 1 \) completes the solution \( (\vartheta, \lambda, \rho, \lambda_1, \lambda_2, \nu) \) of the system (i) to (vi). \( \square \)
References