A PONTRYAGIN MAXIMUM PRINCIPLE FOR INFINITE-DIMENSIONAL PROBLEMS

M. I. KRASTANOVI, N. K. RIBARSKA‡, AND TS. Y. TSACHEV§

Abstract. A basic idea of the classical approach for obtaining necessary optimality conditions in optimal control is to construct suitable “needle-like control variations.” We use this idea to prove the main result of the present paper—a Pontryagin maximum principle for infinite-dimensional optimal control problems with pointwise terminal constraints in arbitrary Banach state space. By refining the classical variational technique we are able to replace the differentiability of the norm of the state space (guaranteed by the strict convexity of its dual norm, which is assumed in the known results) by a separation argument. We also drop another key assumption which is common in the existing literature on infinite-dimensional control problems—that the set of variations (in the state space) of the state trajectory’s endpoint (resulting from the control variations) be finite-codimensional. Instead, we require only that it has nonempty interior in its closed affine hull. As an application of the abstract result we present an illustrative example—an optimal control problem for an age-structured system with pointwise terminal state constraints.

Key words. Pontryagin maximum principle, infinite-dimensional problems, needle-like control variations

AMS subject classifications. 49K20, 49K27, 35F25

DOI. 10.1137/100799009

1. Introduction. Since it was first established about half a century ago, the Pontryagin maximum principle has proved to be a powerful tool for studying optimal control problems. Used to name necessary optimality conditions for the large variety of such problems, this principle takes its specific form in every particular case. While for finite-dimensional systems it is well developed, there are still a number of open problems as far as infinite-dimensional systems are concerned.

The infinite-dimensional version of the Pontryagin maximum principle was first proved by Butkovsky in [5] for systems governed by integral equations and by Kharratishvili [19] for systems governed by ordinary delay equations; both papers were published in 1961. Soon after, in 1964, Yu. V. Egorov constructed in [10] an example showing that the maximum principle does not generally hold for infinite-dimensional systems (the example was first announced in the short note [9] in 1963). In the same paper [10] the author pointed out an additional condition, which, if satisfied, leads to the validity of the maximum principle; in general, it allows the application of a separation theorem. The example of Yu. V. Egorov led to the development of the theory

*Received by the editors June 15, 2010; accepted for publication (in revised form) August 2, 2011; published electronically DATE.

†Department of Biomathematics, Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Acad. G. Bonchev str., bl 8, 1113 Sofia, Bulgaria (krast@math.bas.bg). This author was partially supported by the Bulgarian Ministry of Science and Higher Education National Fund for Science Research under contract DO 02-359/2008.

‡Faculty of Mathematics and Informatics, University of Sofia, James Bourchier Boul. 5, 1126 Sofia, Bulgaria (ribarska@fmi.uni-sofia.bg) and Department of Operations Research, Institute of Mathematics and Informatics, Bulgarian Academy of Sciences. This author was partially supported by the Bulgarian Ministry of Science and Higher Education National Fund for Science Research under contract DO 02-360/2008.

§Department of Operations Research, Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Acad. G. Bonchev str., bl 8, 1113 Sofia, Bulgaria (tsachev@math.bas.bg).
of necessary optimality conditions for infinite-dimensional optimal control problems mainly with no terminal state constraints. An early exception is the paper [8] of A. I. Egorov in which he proved the maximum principle for parabolic and hyperbolic systems with the terminal state constrained by finitely many equalities. About two decades later it was realized that the important feature of such a target set is its finite codimensionality in the state space. In a series of papers Li and Yao [21], Fattorini [11], Fattorini and Frankowska [12], and some others proved the maximum principle for optimal control problems in Banach spaces, assuming finite-codimensionality in the state space of the target set or of some other set closely related to it. Some of the important recent developments in the field of infinite-dimensional control problems with dynamics involving unbounded linear operators are due to Arada, Casas, Raymond, Zidani, and others, who have established necessary (mostly in the form of the Pontryagin maximum principle) and also sufficient optimality conditions (cf., e.g., [1], [2], [6], [7], [24]).

It is the infinite-dimensional technique of Li and Yong, presented in [22], which we use and improve in the present paper. We prove here the maximum principle for an optimal control problem in Banach space with terminal state constraints, which generalizes the main result of Chapter 4 in [22, Theorem 1.6, p. 135]. The generalization consists in dropping one of the two main assumptions, namely, the strict convexity of the dual of the state space. We also relax the other key assumption in the mentioned result from [22]. This second assumption is common in the existing literature on infinite-dimensional control problems. It consists in requiring that the following set be finite-codimensional: the algebraic difference of the set of variations (in the state space) of the state trajectory’s endpoint (resulting from the control variations) and the target set. Instead, we require only that this set have nonempty interior in its closed affine hull. We obtain our result by using technical lemmas from [3] and [20] which allow us to refine the techniques from [22]. Namely, we replace the differentiability of the distance function in the state space (guaranteed by the strict convexity of the dual space) by a separation argument.

Our motivation was to obtain a Pontryagin-type necessary optimality condition for optimal control of age-structured systems, involving pointwise terminal state constraints. The dynamics of such systems is described by distributed state variables (satisfying hyperbolic PDEs of first order and depending on the time and the age) and by aggregate state variables (depending only on the time). This is why unbounded linear operators are not involved in the dynamics of the abstract problem we formulate in section 2. The applications predetermine that the distributed variables satisfy the respective PDEs in a generalized (weak) sense and be taken in such a space that their traces at the terminal time belong to an $L_\infty$ space. This fact necessitated an abstract problem setting in which the dual of the state space is not strictly convex. Another reason to drop the strict convexity of the dual of the state space is the remaining technical assumption that the algebraic difference of the set of variations and the target set (cf. hypothesis (H3) below) has nonempty interior in its closed affine hull in the state space. In the example we present in the last section we show that this algebraic difference contains the negative cone of the state space. Among the $L_p$ spaces, $p \geq 1$, the only one whose negative cone has nonempty interior is $L_\infty$.

Optimal control problems for age-structured systems have been of interest in the last three decades in many areas of application such as economics, demography,

---

3In some models there are aggregate state variables depending not only on the time but also on the age. Here by “distributed state variables” we mean those satisfying the hyperbolic PDE.
epidemiology, illicit drug abuse, and even the social sciences (for a quick overview of the applications we refer the reader to the bibliography in [15]). Many of the published papers present optimality conditions for particular models, usually in the form of a maximum principle of Pontryagin type. A general maximum principle for nonlinear McKendrick-type systems is obtained in [4]. However, a number of extensions of the McKendrick and Gurtin-MacCamy [16] models arose, where the existing optimality conditions were not applicable. A nontrivial extension of the result of Brokate, proved in [15], was able to cope with such situations and therefore found numerous applications, such as in [13], [14], [25], and [26]. The maximum principles of Brokate [4] and of Feichtinger, Tragler, and Veliov [15], as well as their applications to age-structured systems, deal with optimal control problems with free terminal endpoint for the distributed state variables (as well as for the nondistributed variables). To our knowledge, up to now there have been three papers dealing with optimal control problems for age-structured systems, in which pointwise terminal state constraints are present—[23], [17], and [18]. In all three the argument dealing with the terminal constraints is the same, and we think it contains a loophole. This motivated us to prove the Pontryagin maximum principle for the herein discussed abstract problem by applying an entirely different approach. In a follow-up paper we intend to extend this result and apply it to deriving necessary optimality conditions for optimal control problems (in which pointwise terminal state constraints are present) for general age-structured systems. Here we give only an example pointing in this direction.

In section 2 we formulate the abstract problem and state the main result; we present its proof in section 3. In section 4 we present the illustrative example—the optimal control problem (with pointwise terminal state constraints) for an age-structured system. The model is taken from the recent paper [26] on optimal investment in education/training under changing labor demand and supply.

2. Statement of the problem and formulation of the main result. Let us consider the following optimal control problem:

\[
J(u(\cdot), v) := \int_0^T f^0(t, y(t), u(t), v) \, dt \to \min
\]

subject to

\[
\begin{align*}
\dot{y}(t) &= f(t, y(t), u(t), v) \text{ a.e. in } [0, T], \\
y(0) &= \varphi(v) \in X, \\
y(T) &\in S \subseteq X, \\
u(\cdot) &\in \mathcal{U} := \{u(\cdot) : [0, T] \to U, \ u(\cdot) \text{ is strongly measurable}\}, \\
v &\in \mathcal{V} \subseteq Z.
\end{align*}
\]

Here \(X\) and \(Z\) are Banach spaces, and the following hypotheses hold:

(H1) \(S\) is a closed and convex subset of \(X\), \(U\) is a separable metric space, \(\mathcal{V}\) is a bounded closed convex subset of \(Z\), and \(T > 0\) is given.

(H2) The functions \(f : [0, T] \times X \times U \times Z \to X\) and \(f^0 : [0, T] \times X \times U \times Z \to \mathbb{R}\) are strongly measurable for any fixed \((x, u, v) \in X \times U \times Z\). Moreover, \(f\) and \(f^0\) are Fréchet differentiable in \(x\) for any fixed \((t, u, v)\) and in \(v\) for any fixed \((t, x, u)\). At last, the functions \(f(t, \cdot, \cdot, \cdot), f^0(t, \cdot, \cdot, \cdot), f_x^0(t, \cdot, \cdot, \cdot), f^0_x(t, \cdot, \cdot, \cdot), f_u^0(t, \cdot, \cdot, \cdot),\) and \(f^0_u(t, \cdot, \cdot, \cdot)\) are jointly continuous. The function \(\varphi : Z \to X\) is continuously differentiable. There is \(M > 0\) such that \(\|v\| \leq M, \|\varphi'(v)\| \leq M\).
for each $v \in V$, $f'_x(t, x, u, \cdot)$ and $f^0_x(t, x, u, \cdot)$ being Lipschitz continuous with constant $M$ uniformly in $(t, x, u) \in [0, T] \times X \times U$,
\[
\|f^0_x(t, x, u, v)\| \leq M, \quad \|f'_x(t, x, u, v)\| \leq M, \quad \|f'_{x}(t, x, u, v)\| \leq M,
\]
for each $(t, x, u, v) \in [0, T] \times X \times U \times V$.

**Remark 2.1.** In fact, the conditions in $(H_2)$ guarantee that for any $(u(\cdot), v) \in U \times V$ there exists a unique solution $y(\cdot) \in C([0, T]; X)$ of the equation $\dot{y}(t) = f(t, y(t), u(t), v)$.

The following definition is necessary in what follows.

**Definition 2.2.** Let $X$ be a Banach space and $S$ be a subset of $X$. The set $S$ is said to be quasi-solid if its closed convex hull $\overline{S}$ has nonempty interior in its closed affine hull, i.e., if there exists a point $x_0 \in \overline{S}$ such that $\overline{S} \setminus \{x_0\}$ has nonempty interior in $\text{span}(S - x_0)$ (the closed subspace spanned by $S - x_0$).

**Remark 2.3.** Definition 2.2 is closely related to Definition 1.5 ([22, p. 134]), for a set $S$ to be finite-codimensional. The latter says that $S$ is finite-codimensional in $X$ if there exists a point $x_0 \in \overline{S}$ such that the closed subspace spanned by $\{z - x_0 : z \in S\}$ is a finite-codimensional subspace of $X$ and $\overline{S} \setminus \{x_0\}$ has nonempty interior in this subspace. The only difference between the two definitions is that the requirement for finite-codimensionality of $\text{span}(S - x_0)$ is dropped in our Definition 2.2.

**Remark 2.4.** For the case of a Hilbert space $X$, Definition 1.5 in [22] contains the definition of "set of finite codimension" given on p. 160 in [11]: It is said that the set $C$ has finite codimension if there exists a closed subspace $Y$ of $X$ of finite codimension such that $C_Y := \text{pr}_Y(\overline{C})$ has nonempty interior in $Y$, where $\text{pr}_Y$ denotes the orthogonal projection from $X$ to $Y$.

Indeed, let $C$ be a closed convex and balanced subset of the Hilbert space $X$, where $X = X_1 \oplus X_2$ with $\dim X_2 < \infty$. We are going to show that if $\text{relint} X_1 \left( \text{pr}_{X_1} C \right) \neq \emptyset$, then $\text{relint}_Y C \neq \emptyset$, where $Y = \text{span}(C)$. To do this we are going to prove that $C$ is absorbing in $Y$. Assume the contrary, i.e., $Y \setminus \left( \bigcup_{n=1}^{\infty} nC \right) \neq \emptyset$. Take $x \in Y \setminus \left( \bigcup_{n=1}^{\infty} nC \right)$. Then $x/n \notin C$ for each positive integer $n$. We next apply the separation theorem in $Y$: there exists $f_n$ with $\|f_n\| = 1$ such that $\langle f_n, x/n \rangle \leq \langle f_n, c \rangle$ for each $c \in C$. We can extend $f_n$ to be from $X^*$ and such that $\langle f_n, z \rangle = 0$ for each $z \in Z$, where $X = Y \oplus Z$. Since $-1/n\|x\| \leq \langle f_n, x/n \rangle \leq \langle f_n, c \rangle$ for each $c \in C$ and $\|f_n\|_{X^*} = 1$ for each positive integer $n$, we can extract a subsequence $\{f_{n_k}\}_{k=1}^{\infty}$ such that $f_{n_k} \rightharpoonup \tilde{f} \in X^*$ as $k \to \infty$, and according to Lemma 5.6 from [11], we have that $\tilde{f} \neq 0$. Also, $0 \leq \langle \tilde{f}, c \rangle$ for each $c \in C$. Since $C$ is balanced, $0 = \langle \tilde{f}, c \rangle$ for each $c \in C$. From $0 = \langle f_n, z \rangle$ for each $z \in Z$, it follows that $0 = \langle \tilde{f}, z \rangle$ for each $z \in Z$. Hence, $\langle \tilde{f}, y \rangle \neq 0$ for some $y \in Y$. This and the equality $\langle \tilde{f}, c \rangle = 0$ for each $c \in C$ imply that $C$ is contained in a proper subspace of $Y$. This contradicts $Y = \text{span}(C)$. Hence, $C$ is absorbing in $Y$. Applying the Baire category theorem we obtain that $C$ has a nonempty interior in $Y$.

In what follows we denote by $(\bar{x}, \bar{u}, \bar{v})$ an optimal triple for the problem (2.1)–(2.2), $(\bar{u}, \bar{v})$ being optimal controls and $\bar{x}$ the corresponding trajectory.
We set \( \Xi := \bigcup_{k=1}^{\infty} \Xi_k \), where

\[
\Xi_k := \left\{ \{u_i(\cdot)\}_{i=1}^{k}, \{v_i\}_{i=1}^{k}, \{\lambda_i\}_{i=1}^{k} : u_i(\cdot) \in \mathcal{U}, v_i \in \mathcal{V}, \lambda_i \geq 0, i = 1, \ldots, k, \sum_{i=1}^{k} \lambda_i = 1 \right\}
\]

for each positive integer \( k \), and define

\[
\mathcal{P} := \left\{ \zeta(T) \in X : \text{there exist } \theta = \{\{u_i(\cdot)\}_{i=1}^{k}, \{v_i\}_{i=1}^{k}, \{\lambda_i\}_{i=1}^{k}\} \in \Xi,
\begin{align*}
&\text{such that } \zeta(t) = \varphi'(\bar{v}) \circ \left( \sum_{i=1}^{k} \lambda_i v_i - \bar{v} \right) + \int_{0}^{t} f_x'(s, \bar{x}(s), \bar{u}(s), \bar{v}).\zeta(s) \, ds \\
&\quad + \sum_{i=1}^{k} \lambda_i \int_{0}^{t} (f(s, \bar{x}(s), u_i(s), \bar{v}) - f(s, \bar{x}(s), \bar{u}(s), \bar{v})) \, ds \\
&\quad + \int_{0}^{t} f_x'(s, \bar{x}(s), \bar{u}(s), \bar{v}) \, ds \circ \left( \sum_{i=1}^{k} \lambda_i v_i - \bar{v} \right), t \in [0, T] \right\}
\end{align*}
\]

The next assumption is crucial in the proof of the main result.

(H3) The subset \( \mathcal{P} - \mathcal{S} \) of \( X \) is quasi-solid.

We are now able to formulate the main result.

**Theorem 2.5.** Let the assumptions (H1), (H2), and (H3) hold true. If \((\bar{x}, \bar{u}, \bar{v})\) is an optimal triple of the problem (2.1)–(2.2), then there exists a nontrivial pair \((\psi_0, \psi(\cdot)) \in \mathbb{R}^1 \times C([0, T], X^*)\) such that \( \psi_0 \leq 0 \),

\[
\dot{\psi}(t) = - (f_x'(t, \bar{x}(t), \bar{u}(t), \bar{v}))^* \psi(t) - \psi^0 f_x'(t, \bar{x}(t), \bar{u}(t), \bar{v}),
\]

and

\[
H(t, \bar{x}(t), \bar{u}(t), \bar{v}, \psi_0, \psi(t)) = \max_{u \in \mathcal{U}} H(t, \bar{x}(t), u, \bar{v}, \psi_0, \psi(t)) \text{ a.e. in } [0, T],
\]

\[
\left\langle \psi(T), x_1 - \bar{x}(T) \right\rangle \leq 0 \quad \text{for each } x_1 \in \mathcal{S},
\]

\[
\left\langle (\varphi'(\bar{v}))^* \psi(0) + \int_{0}^{T} H_{\psi'}(t, \bar{x}(t), \bar{u}(t), \bar{v}) \, dt, (v - \bar{v}) \right\rangle \leq 0
\]

for each \( v \in \mathcal{V} \), where

\[
H(t, x, u, v, \psi_0, \psi) = \left\langle \psi, f(t, x, u, v) \right\rangle + \psi_0 f_x'(t, x, u, v).
\]

3. **Proof of the main result.** The proof is divided into several steps.

**Step 1. Some continuity properties of the trajectory and of the cost functional.**

**Definition 3.1.** We set

\[
d(\langle u(\cdot), v \rangle, \langle \bar{u}(\cdot), \bar{v} \rangle) := \sqrt{d^2(\bar{u}(\cdot), u(\cdot)) + \| \bar{v} - v \|_2^2}
\]

to be the distance between two admissible control pairs \( (u(\cdot), v) \) and \( (\bar{u}(\cdot), \bar{v}) \) where,

\[
d(\bar{u}(\cdot), u(\cdot)) := \text{meas } \{ t \in [0, T] : \bar{u}(\cdot) \neq u(t) \}.
\]
The following lemma is very similar to Lemma 4.1 of [22, p. 151], and can be proved in the same way.

**Lemma 3.2.** The hypotheses (H1) and (H2) imply the existence of constants \( C_1 > 0 \) and \( C_2 > 0 \) such that for all \( u(\cdot) \) and \( \bar{u}(\cdot) \) of \( \mathcal{U} \), and for all \( v \) and \( \bar{v} \) of \( \mathcal{V} \), we have

\[
\sup_{t \in [0,T]} \| x(t, u(\cdot), v) - x(t, \bar{u}(\cdot), \bar{v}) \| \leq C_1 \delta((u(\cdot), v), (\bar{u}(\cdot), \bar{v}))
\]

and

\[
| J(u(\cdot), v) - J(\bar{u}(\cdot), \bar{v}) | \leq C_2 \delta((u(\cdot), v), (\bar{u}(\cdot), \bar{v})).
\]

Lemma 3.2 directly implies the following corollary.

**Corollary 3.3.** The functional \( J(u(\cdot), v) \) is Lipschitz continuous in \( (u(\cdot), v) \) with respect to the distance function \( \delta \).

**Step 2. Applying Ekeland’s variational principle.** Without loss of generality we can assume that \( J(\bar{u}(\cdot), \bar{v}) = 0 \). For each \( \varepsilon > 0 \) we define the penalty functional

\[
J_\varepsilon(u(\cdot), v) := \sqrt{d_S^2(x(T, u(\cdot), v)) + (J(u(\cdot), v) + \varepsilon)^+}^2,
\]

where

\[
a^+ := \max\{a, 0\} \quad \text{and} \quad d_S(x) := \inf_{s \in S} \| x - s \|.
\]

This definition implies that for each \( \varepsilon > 0 \), we have \( J_\varepsilon(u(\cdot), v) > 0 \) (since we have assumed that \( J(\bar{u}(\cdot), \bar{v}) = 0 \), i.e., \( J(u(\cdot), v) \geq 0 \) for all admissible \( (u(\cdot), v) \in \mathcal{U} \times \mathcal{V} \)). Moreover,

\[
J_\varepsilon(\bar{u}(\cdot), \bar{v}) = \varepsilon \leq \inf \{ J_\varepsilon(u(\cdot), v) \colon (u(\cdot), v) \in \mathcal{U} \times \mathcal{V} \} + \varepsilon.
\]

From the Ekeland variational principle there exists \((u_\varepsilon(\cdot), v_\varepsilon) \in \mathcal{U} \times \mathcal{V}\) such that

\[
\tilde{d}((u_\varepsilon(\cdot), v_\varepsilon), (\bar{u}(\cdot), \bar{v})) \leq \sqrt{\varepsilon} \quad \text{and} \quad J_\varepsilon(u(\cdot), v) + \sqrt{\varepsilon} \tilde{d}((u_\varepsilon(\cdot), v_\varepsilon), (u(\cdot), v)) \geq J_\varepsilon(u_\varepsilon(\cdot), v_\varepsilon) \quad \text{for each} \quad (u(\cdot), v) \in \mathcal{U} \times \mathcal{V}.
\]

**Step 3. Families of control variations \((u^\theta_{\varepsilon, \rho}(\cdot), v^\theta_{\varepsilon, \rho})\).** In this step we recall the concept of diffuse control variations developed, e.g., in [12], [22]. But in contrast to these references we use the extension proposed in [20], leading later to the convexity of the corresponding state variations.

Let \( \varepsilon > 0 \) be given. Let us fix \( \theta := ((u_1(\cdot), v_1, \lambda_1), (u_2(\cdot), v_2, \lambda_2), \ldots, (u_k(\cdot), v_k, \lambda_k)) \) belonging to \( \Xi \), i.e.,

\[
u_i(\cdot) \in \mathcal{U}, \quad v_i \in \mathcal{V}, \quad \lambda_i \geq 0, \quad i = 1, \ldots, k, \quad \text{with} \quad \sum_{i=1}^k \lambda_i = 1.
\]

We set

\[
\mathbb{R}^1 \times X \ni h_\varepsilon(s) := \begin{pmatrix}
    f^0(s, x_\varepsilon(s), u_i(s), v_\varepsilon) - f^0(s, x_\varepsilon(s), u_\varepsilon(s), v_\varepsilon) \\
    f(s, x_\varepsilon(s), u_i(s), v_\varepsilon) - f(s, x_\varepsilon(s), u_\varepsilon(s), v_\varepsilon)
\end{pmatrix}
\]
for each \( i = 1, \ldots, k \). As is seen, \( h_i, i = 1, \ldots, k \), do not depend on \( t \). It follows from the assumptions that

\[
\max_{0 \leq s \leq t} \| x(s, u_\varepsilon(\cdot), v_\varepsilon) \| \leq \text{const}, \quad |f^0(s, x(s, u_\varepsilon(\cdot), v_\varepsilon), u_1(s), v_1)| \leq M
\]

and \( |f^0(s, x(s, u_\varepsilon(\cdot), v_\varepsilon), u_1(s), v_1)| \leq M \) on \([0, T] \times \mathcal{V}\). The same holds true for \( \| f(\cdot, \cdot, \cdot, \cdot) \| \). Hence, we can apply Corollary 3.9 from [22, p. 144] and Lemma 5.5 from [20, p. 358]: For \( h_i : [0, T] \to \mathbb{R}^1 \times X \) defined above and for \( \{ \lambda_i \}_{i=1}^k \) there exist measurable subsets \( \{ F_{\rho,i} \}_{i=1}^k \) of \([0, T] \) such that \( F_{\rho,i} \cap F_{\rho,j} = \emptyset \) for \( i \neq j \), \( \text{meas} \left( \bigcup_{i=1}^k F_{\rho,i} \right) = \rho T \), and

\[
(3.1) \quad \rho \sum_{j=1}^k \lambda_j \int_0^t h_j(s) \, ds = \sum_{j=1}^k \int_{F_{\rho,j} \cap [0,t]} h_j(s) \, ds + o(\rho),
\]

where \( o(\rho)/\rho \) tends to zero as \( \rho \to 0 \) uniformly with respect to \( \theta \in \Xi \) and to \( t \in [0, T] \).

We introduce the control

\[
u_{\varepsilon, \rho}^\theta = \begin{cases} u_i(t) & \text{for } t \in F_{\rho,i}, \quad i = 1, 2, \ldots, k, \\ u_\varepsilon(t) & \text{for } t \in [0, T] \setminus \bigcup_{i=1}^k F_{\rho,i} \end{cases}
\]

and obtain that

\[
(3.2) \quad \sum_{j=1}^k \int_{F_{\rho,j} \cap [0,t]} h_j(s) \, ds = \int_0^t h_{\rho}^0(s) \, ds,
\]

where

\[
h_{\rho}^\theta(s) := \begin{pmatrix} h_{\rho}^{\theta_1}(s) \\ h_{\rho}^{\theta_2}(s) \end{pmatrix} := \begin{pmatrix} f^0(s, x_\varepsilon(s), u_\varepsilon(\cdot), v_\varepsilon) - f^0(s, x_\varepsilon(s), u_\varepsilon(s), v_\varepsilon) \\ f(s, x_\varepsilon(s), u_\varepsilon(\cdot), v_\varepsilon) - f(s, x_\varepsilon(s), u_\varepsilon(s), v_\varepsilon) \end{pmatrix}.
\]

Taking into account (3.1) and (3.2), we obtain that

\[
(3.3) \quad \rho \sum_{j=1}^k \lambda_j \int_0^t h_j(s) \, ds = \int_0^t h_{\rho}^0(s) \, ds + o(\rho),
\]

where \( o(\rho)/\rho \) tends to zero uniformly with respect to \( \theta \in \Xi \) and to \( t \in [0, T] \) as \( \rho \to 0 \).

**Step 4. State variations corresponding to the control variations.** The goal of this step is to prove a technical lemma which gives a presentation of the state variation as a solution of a linear differential equation, as well as a respective result for the cost functional.

We set \( v_{\varepsilon} := \sum_{j=1}^k \lambda_j v_j, \quad v_{\varepsilon, \rho}^\theta := v_\varepsilon + \rho (v_{\varepsilon} - v_\varepsilon) \) and denote \( x_{\varepsilon, \rho}^\theta(\cdot) := x(\cdot, u_{\varepsilon, \rho}^\theta, v_{\varepsilon, \rho}^\theta) \). Next we establish Taylor-like expansions with respect to \( \rho \) at \( \rho = 0 \).

**Lemma 3.4.** We claim that

\[
\sup_{\theta \in \Xi} \sup_{t \in [0, T]} \left( \| x_{\varepsilon, \rho}^\theta(t) - x_\varepsilon(t) - \rho z_{\varepsilon}^\theta(t) \|_{\mathcal{X}} \right) = o(\rho)
\]

and

\[
\sup_{\theta \in \Xi} | J(u_{\varepsilon, \rho}^\theta(\cdot), v_{\varepsilon, \rho}^\theta) - J(u_\varepsilon(\cdot), v_\varepsilon) - \rho z_{\varepsilon}^\theta(0) | = o(\rho),
\]
where \( z_\varepsilon^\theta(\cdot) \) is the solution of the following differential equation:

\[
\begin{align*}
   z_\varepsilon^\theta(\cdot) &= \varphi'(v_\varepsilon)(v^\theta - v_\varepsilon) + \int_0^t f'_x(s, x_\varepsilon(s), u_\varepsilon(s), v_\varepsilon) \, z_\varepsilon^\theta(s) \, ds \\
   &\quad + \sum_{j=1}^k \lambda_j \int_0^t (f(s, x_\varepsilon(s), u_j(s), v_\varepsilon) - f(s, x_\varepsilon(s), u_\varepsilon(s), v_\varepsilon)) \, ds \\
   &\quad + \int_0^t f'_x(s, x_\varepsilon(s), u_\varepsilon(s), v_\varepsilon) (v^\theta - v_\varepsilon) \, ds
\end{align*}
\]

and

\[
\begin{align*}
   z_\varepsilon^{\theta,0} := \sum_{i=1}^k \lambda_i \int_0^T (f_0^0(s, x_\varepsilon(s), u_i(s), v_\varepsilon) - f_0^0(s, x_\varepsilon(s), u_\varepsilon(s), v_\varepsilon)) \, ds \\
   &\quad + \int_0^T f_0'(s, x_\varepsilon(s), u_\varepsilon(s), v_\varepsilon) z_\varepsilon^\theta(s) \, ds + \int_0^T f_0'(s, x_\varepsilon(s), u_\varepsilon(s), v_\varepsilon) (v^\theta - v_\varepsilon) \, ds.
\end{align*}
\]

Proof. Denote

\[
\begin{align*}
   z_{\varepsilon,\rho}^\theta(t) := \frac{1}{\rho} \left( x(t, u_{\varepsilon,\rho}^\theta, v_{\varepsilon,\rho}^\theta) - x_\varepsilon(t, u_\varepsilon, v_\varepsilon) \right).
\end{align*}
\]

We have

\[
\begin{align*}
   z_{\varepsilon,\rho}^\theta(t) \\
   := \frac{1}{\rho} \left( \varphi'(v_\varepsilon^\theta) + \int_0^t f(s, x_{\varepsilon,\rho}^\theta(s), u_{\varepsilon,\rho}^\theta(s), v_{\varepsilon,\rho}^\theta(s)) \, ds - \varphi(v_\varepsilon) - \int_0^t f(s, x_\varepsilon(s), u_\varepsilon(s), v_\varepsilon) \, ds \right) \\
   = \varphi'(v_\varepsilon)(v^\theta - v_\varepsilon) + \frac{o(\rho)}{\rho} \\
   &\quad + \frac{1}{\rho} \int_0^t \left( f(s, x_{\varepsilon,\rho}^\theta(s), u_{\varepsilon,\rho}^\theta(s), v_{\varepsilon,\rho}^\theta(s)) - f(s, x_\varepsilon(s), u_\varepsilon(s), v_\varepsilon) \right) \, ds \\
   &\quad + \frac{1}{\rho} \int_0^t \left( f(s, x_\varepsilon(s), u_\varepsilon(s), v_\varepsilon) - f(s, x_\varepsilon(s), u_{\varepsilon,\rho}^\theta(s), v_{\varepsilon,\rho}^\theta(s)) \right) \, ds \\
   &\quad + \frac{1}{\rho} \int_0^t \left( f(s, x_\varepsilon(s), u_{\varepsilon,\rho}^\theta(s), v_\varepsilon) - f(s, x_\varepsilon(s), u_{\varepsilon,\rho}^\theta(s), v_\varepsilon) \right) \, ds,
\end{align*}
\]

where

\[
\frac{o(\rho)}{\rho} := \left\| \int_0^1 (\varphi'(v_\varepsilon + \sigma \rho (v^\theta - v_\varepsilon)) - \varphi'(v_\varepsilon)) (v^\theta - v_\varepsilon) \, d\sigma \right\|_X
\]

\[
\leq 2M \int_0^1 \| \varphi'(v_\varepsilon + \sigma \rho (v^\theta - v_\varepsilon)) - \varphi'(v_\varepsilon) \|_{L(Z,X)} \, d\sigma.
\]

Because \( \varphi'(\cdot) \) is assumed continuous, the right-hand side tends to zero as \( \rho \to 0 \) uniformly with respect to \( \theta \in \Xi. \)
Applying (3.3) we obtain that
\[
z_{\varepsilon,\rho}(t) := \varphi'(v_{c})(v_{\theta} - v_{\varepsilon}) + \int_{0}^{t} \left( f'_{x}(s, \sigma_{x_{\varepsilon,\rho}}(s), u_{x_{\varepsilon,\rho}}(s), v_{x_{\varepsilon,\rho}}(s)) + (1 - \sigma) x_{\varepsilon}(s), u_{x_{\varepsilon,\rho}}(s), v_{x_{\varepsilon,\rho}}(s) \right) ds + \frac{o(\rho)}{\rho}
\]

Hence we have
\[
z_{\varepsilon,\rho}(t) - z_{\varepsilon}(t)
= \int_{0}^{t} \left( f'_{x}(s, \sigma_{x_{\varepsilon,\rho}}(s), u_{x_{\varepsilon,\rho}}(s), v_{x_{\varepsilon,\rho}}(s)) - f'_{x}(s, \sigma_{x_{\varepsilon,\rho}}(s), u_{x_{\varepsilon,\rho}}(s), v_{x_{\varepsilon,\rho}}(s)) \right) ds
+ \int_{0}^{t} \left( f'_{x}(s, \sigma_{x_{\varepsilon,\rho}}(s), u_{x_{\varepsilon,\rho}}(s), v_{x_{\varepsilon,\rho}}(s)) - f'_{x}(s, \sigma_{x_{\varepsilon,\rho}}(s), u_{x_{\varepsilon,\rho}}(s), v_{x_{\varepsilon,\rho}}(s)) \right) ds
+ \int_{0}^{t} \left( f'_{x}(s, \sigma_{x_{\varepsilon,\rho}}(s), u_{x_{\varepsilon,\rho}}(s), v_{x_{\varepsilon,\rho}}(s)) - f'_{x}(s, \sigma_{x_{\varepsilon,\rho}}(s), u_{x_{\varepsilon,\rho}}(s), v_{x_{\varepsilon,\rho}}(s)) \right) ds
+ \int_{0}^{t} \left( f'_{x}(s, \sigma_{x_{\varepsilon,\rho}}(s), u_{x_{\varepsilon,\rho}}(s), v_{x_{\varepsilon,\rho}}(s)) - f'_{x}(s, \sigma_{x_{\varepsilon,\rho}}(s), u_{x_{\varepsilon,\rho}}(s), v_{x_{\varepsilon,\rho}}(s)) \right) ds.

The definition of \( z_{\varepsilon}(\cdot) \) implies
\[
\| z_{\varepsilon}(t) \|_{X} \leq \| \varphi'(v_{c}) \|_{L(Z,X)} \text{ diam } (V) + M \int_{0}^{t} \| z_{\varepsilon}(s) \|_{X} ds + 2MT,
\]
and the Gronwall inequality yields
\[
\| z_{\varepsilon}(t) \|_{X} \leq e^{MT} \left( 2M^{2} + 2M(M + 1)T \right)
\]
for each \( \theta \in \Xi \) and each \( t \in [0, T] \). Therefore, using the Lipschitz continuity in \( v \) of \( f'_{x} \) uniformly in \( (t, y, u) \), the first term in the right-hand side of the last equality can be estimated by \( \rho e^{MT} (2M^{2} + 2M(M + 1)T) \text{ diam}_{Z} V \) and \( \text{ diam}_{Z} V \leq 2M \). The second term can be written as follows:
\[
\int_{[0,t] \cap F_{\rho}} \left( \int_{0}^{1} f'_{x}(s, \sigma_{x_{\varepsilon,\rho}}(s), u_{x_{\varepsilon,\rho}}(s), v_{x_{\varepsilon,\rho}}(s)) - f'_{x}(s, \sigma_{x_{\varepsilon,\rho}}(s), u_{x_{\varepsilon,\rho}}(s), v_{x_{\varepsilon,\rho}}(s)) \right) ds
\]
Lemma 3.2 implies

\[ z(t) \leq C \| z(t) \|_X \]

uniformly with respect to \( t \). Applying the Gronwall inequality, we obtain

\[ z(t) \leq e^{Ct} \]

uniformly with respect to \( t \). This completes the proof. □

Thus the first assertion of the lemma is proved. The second assertion can be proved analogously. This completes the proof. □
Step 5. The limit case as $\rho$ tends to zero. Using the result from the previous step and the Ekeland variational principle, we derive a nontrivial inequality involving the state variations.

For arbitrary

$$\theta = ((u_1(\cdot), v_1, \lambda_1), (u_2(\cdot), v_2, \lambda_2), \ldots, (u_k(\cdot), v_k, \lambda_k)) \in \Xi$$

we obtain according to the Ekeland variational principle that for any $\rho > 0$

$$\frac{1}{\rho} \left( J_\varepsilon(u_{\varepsilon, \rho}^\theta(\cdot), v_{\varepsilon, \rho}^\theta) - J_\varepsilon(u_\varepsilon(\cdot), v_\varepsilon) \right)$$

$$\geq -\frac{\sqrt{\varepsilon}}{\rho} d \left( ((u_{\varepsilon, \rho}^\theta(\cdot), v_{\varepsilon, \rho}^\theta), (u_\varepsilon(\cdot), v_\varepsilon)) \right)$$

$$= -\frac{\sqrt{\varepsilon}}{\rho} \sqrt{\left( \text{meas} \left\{ t \in [0, T] : u_{\varepsilon, \rho}^\theta(t) \neq u_\varepsilon(t) \right\} \right)^2 + \rho^2 \| v^\theta - v_\varepsilon \|^2}$$

$$\geq -\frac{\sqrt{\varepsilon}}{\rho} \sqrt{\rho^2 T^2 + \rho^2 \| v^\theta - v_\varepsilon \|^2} = -\sqrt{\varepsilon} \sqrt{T^2 + \| v^\theta - v_\varepsilon \|^2}.$$

On the other hand,

$$\frac{1}{\rho} \left( J_\varepsilon(u_{\varepsilon, \rho}^\theta(\cdot), v_{\varepsilon, \rho}^\theta) - J_\varepsilon(u_\varepsilon(\cdot), v_\varepsilon) \right)$$

$$= \frac{1}{\rho} \frac{d_S^2(x(T, u_{\varepsilon, \rho}^\theta(\cdot), v_{\varepsilon, \rho}^\theta)) - d_S^2(x(T, u_\varepsilon(\cdot), v_\varepsilon))}{d_S(x(T, u_\varepsilon(\cdot), v_\varepsilon)) + d_S(x(T, u_\varepsilon(\cdot), v_\varepsilon))} + \frac{1}{\rho} \frac{\left( (J(u_{\varepsilon, \rho}^\theta(\cdot), v_{\varepsilon, \rho}^\theta) + \varepsilon)^+ - (J(u_\varepsilon(\cdot), v_\varepsilon) + \varepsilon)^+ \right)^2}{d_S(x(T, u_\varepsilon(\cdot), v_\varepsilon)) + d_S(x(T, u_\varepsilon(\cdot), v_\varepsilon))}.$$

Moreover,

$$\lim_{\rho \downarrow 0} \frac{1}{\rho} \left( d_S^2(x(T, u_{\varepsilon, \rho}^\theta(\cdot), v_{\varepsilon, \rho}^\theta)) - d_S^2(x(T, u_\varepsilon(\cdot), v_\varepsilon)) \right)$$

$$= \lim_{\rho \downarrow 0} \frac{1}{\rho} \left( d_S(x(T, u_{\varepsilon, \rho}^\theta(\cdot), v_{\varepsilon, \rho}^\theta)) - d_S(x(T, u_\varepsilon(\cdot), v_\varepsilon)) \right),$$

$$(d_S(x(T, u_{\varepsilon, \rho}^\theta(\cdot), v_{\varepsilon, \rho}^\theta)) + d_S(x(T, u_\varepsilon(\cdot), v_\varepsilon)))$$

$$= 2d_S(x(T, u_\varepsilon(\cdot), v_\varepsilon)) \lim_{\rho \downarrow 0} \frac{1}{\rho} \left( d_S(x(T, u_\varepsilon(\cdot), v_\varepsilon) + \rho z_\varepsilon(T) + o(\rho)) - d_S(x(T, u_\varepsilon(\cdot), v_\varepsilon)) \right)$$

$$= 2d_S(x(T, u_\varepsilon(\cdot), v_\varepsilon)) \left( \lim_{\rho \downarrow 0} \frac{1}{\rho} \left( d_S(x(T, u_\varepsilon(\cdot), v_\varepsilon) + \rho z_\varepsilon(T)) - d_S(x(T, u_\varepsilon(\cdot), v_\varepsilon)) \right) + \lim_{\rho \downarrow 0} \frac{1}{\rho} \left( d_S(x(T, u_\varepsilon(\cdot), v_\varepsilon) + \rho z_\varepsilon(T) + o(\rho)) - d_S(x(T, u_\varepsilon(\cdot), v_\varepsilon) + \rho z_\varepsilon(T)) \right) \right)$$

$$= 2d_S(x(T, u_\varepsilon(\cdot), v_\varepsilon)) \lim_{\rho \downarrow 0} \frac{1}{\rho} \left( d_S(x(T, u_\varepsilon(\cdot), v_\varepsilon) + \rho z_\varepsilon(T)) - d_S(x(T, u_\varepsilon(\cdot), v_\varepsilon)) \right).$$
Next, we obtain that
\[
\lim_{\rho \to 0} \frac{1}{\rho} \left( (J(u_{\epsilon,\rho}^\theta(\cdot), v_{\epsilon,\rho}^\theta)) + \epsilon \right)^+ - \left( (J(u_{\epsilon}(\cdot), v_{\epsilon}) + \epsilon)^+ \right)^2
\]
\[
= \lim_{\rho \to 0} \frac{1}{\rho} \left( (J(u_{\epsilon,\rho}^\theta(\cdot), v_{\epsilon,\rho}^\theta)) + \epsilon \right)^+ + (J(u_{\epsilon}(\cdot), v_{\epsilon}) + \epsilon)^+
\]
\[
= \lim_{\rho \to 0} \frac{1}{\rho} \left( (J(u_{\epsilon,\rho}^\theta(\cdot), v_{\epsilon,\rho}^\theta)) + \epsilon \right)^+ - (J(u_{\epsilon}(\cdot), v_{\epsilon}) + \epsilon)^+
\]
\[
= 2 (J(u_{\epsilon}(\cdot), v_{\epsilon}) + \epsilon)^+ z_{\epsilon}^\theta,0.
\]
The last equality is evident whenever \((J(u_{\epsilon}(\cdot), v_{\epsilon}) + \epsilon)^+ = 0\). If this is not the case,
\[
\lim_{\rho \to 0} \frac{1}{\rho} \left( (J(u_{\epsilon,\rho}^\theta(\cdot), v_{\epsilon,\rho}^\theta)) + \epsilon \right)^+ - (J(u_{\epsilon}(\cdot), v_{\epsilon}) + \epsilon)^+
\]
\[
= \lim_{\rho \to 0} \frac{1}{\rho} (J(u_{\epsilon,\rho}^\theta(\cdot), v_{\epsilon,\rho}^\theta)) + \epsilon - (J(u_{\epsilon}(\cdot), v_{\epsilon}) - \epsilon) = z_{\epsilon}^\theta,0.
\]
Taking \(\rho\) to tend to 0 in (3.4) and (3.5), we obtain that
\[
\lim_{\rho \to 0} \frac{1}{\rho} (J(u_{\epsilon,\rho}^\theta(\cdot), v_{\epsilon,\rho}^\theta) - J(u_{\epsilon}(\cdot), v_{\epsilon}))
\]
\[
= \frac{d_S(x(T, u_{\epsilon}(\cdot), v_{\epsilon}))}{J_{\epsilon}(u_{\epsilon}(\cdot), v_{\epsilon})} \lim_{\rho \to 0} \frac{1}{\rho} (d_S(x(T, u_{\epsilon}(\cdot), v_{\epsilon}) + \rho z_{\epsilon}^\theta(T)) - d_S(x(T, u_{\epsilon}(\cdot), v_{\epsilon}))
\]
\[
+ \frac{(J(u_{\epsilon}(\cdot), v_{\epsilon}) + \epsilon)^+ z_{\epsilon}^\theta,0}{J_{\epsilon}(u_{\epsilon}(\cdot), v_{\epsilon})} \geq -\sqrt{\epsilon} \sqrt{T^2 + \|v^\theta - v_{\epsilon}\|^2}.
\]
Hence (here \(\epsilon\) is fixed)
\[
\frac{d_S(x(T, u_{\epsilon}(\cdot), v_{\epsilon}))}{J_{\epsilon}(u_{\epsilon}(\cdot), v_{\epsilon})} d_\theta^0 (x(T, u_{\epsilon}(\cdot), v_{\epsilon}); z_{\epsilon}^\theta(T)) + \frac{(J(u_{\epsilon}(\cdot), v_{\epsilon}) + \epsilon)^+ z_{\epsilon}^\theta,0}{J_{\epsilon}(u_{\epsilon}(\cdot), v_{\epsilon})} \geq -\sqrt{\epsilon} \sqrt{T^2 + 4M^2},
\]
and this is true for every \(\theta\) belonging to \(\Xi\) (here we have denoted by \(d_\theta^0(y; z)\) the directional derivative of the function \(d_S\) at the point \(y\) in direction \(z\)).

**Step 6. Convexity of the state variations and separation.** This step is crucial for the proof and deviates substantially from the framework of [22, Chap. 4]. It provides a nontrivial functional separating the “augmented variations.”

**Lemma 3.5.** Let \(\epsilon > 0\) be given. Then the set \(\{(z_{\epsilon}^\theta(T), z_{\epsilon}^\theta,0) : \theta \in \Xi\}\) is convex in \(X \times \mathbb{R}\).

**Proof.** We have to show that for arbitrary elements
\[
\theta_i = \left\{u_{i,j}^j \right\}_{j=1}^{k_i}, \left\{v_{i,j}^j \right\}_{j=1}^{k_i}, \left\{\Lambda_{i,j}^j \right\}_{j=1}^{k_i} \in \Xi, \quad i = 1, \ldots, m,
\]
and arbitrary nonnegative \(\mu_i, i = 1, \ldots, m,\) with \(\sum_{i=1}^{m} \mu_i = 1\) there exists \(\theta \in \Xi\) such that
\[
z_{\epsilon}^\theta(t) = \sum_{i=1}^{m} \mu_i z_{\epsilon}^\theta_i(t), \quad t \in [0, T], \quad \text{and} \quad z_{\epsilon}^\theta,0 = \sum_{i=1}^{m} \mu_i z_{\epsilon}^\theta,0_i.
We set
\[ \theta = \left\{ \{ u^i_j(\cdot) \}_{i=1}^{k_i}, \{ v^j_i(\cdot) \}_{j=1}^{k_j}, \{ \mu_i \gamma^i_j \}_{i=1}^{k_i} \right\}. \]

Clearly, \( \theta \in \Xi \), and it can be directly checked that the uniqueness of the solution of the linear ODE implies the required conclusion. \( \square \)

By setting \( L := \sqrt{T^2 + 4M^2} \), \( y_\varepsilon(T) := y(T, u_\varepsilon(\cdot), v_\varepsilon) \),
\[ a_\varepsilon := \frac{d_S(x(T, u_\varepsilon(\cdot), v_\varepsilon))}{J_\varepsilon(u_\varepsilon(\cdot), v_\varepsilon)}, \quad b_\varepsilon := \frac{(J(u_\varepsilon(\cdot), v_\varepsilon) + \varepsilon)^+}{J_\varepsilon(u_\varepsilon(\cdot), v_\varepsilon)}, \]
we can write (3.6) as follows:
\[ a_\varepsilon d_S^0(x_\varepsilon(T); z_\varepsilon^0(T)) + b_\varepsilon z_\varepsilon^0 \geq -L\sqrt{\varepsilon}. \]

We note that \( a_\varepsilon^2 + b_\varepsilon^2 = 1. \)

We next define the two sets in \( X \times \mathbb{R}^1 \times \mathbb{R}^1 \):
\[ A_\varepsilon := \{(z, z^0, r) : a_\varepsilon d_S^0(x_\varepsilon(T); z) + b_\varepsilon z^0 < r \} \]
and
\[ B_\varepsilon := \{(z, z^0, r) : \text{there exists } \theta \in \Xi \text{ such that } (z, z^0, r) = (z_\varepsilon^0(T), z_\varepsilon^0, -L\sqrt{\varepsilon}) \}. \]

The Lipschitz continuity of the function \( d_S^0(y_\varepsilon(T), \cdot) \) implies that the set \( A_\varepsilon \) is open. Also, it is nonempty (\( d_S^0(y_\varepsilon(T); z) \) is finite) and convex (\( (d_S^0(y_\varepsilon(T); \cdot) \) is convex). The set \( B_\varepsilon \) is convex because of Lemma 3.5. Then, according to the separation theorem, there exists \( 0 \neq (\xi_\varepsilon, p_\varepsilon, q_\varepsilon) \in X^* \times \mathbb{R}^1 \times \mathbb{R}^1 \) such that
\[ \langle \xi_\varepsilon, z \rangle_{(X^*, X)} + p_\varepsilon z^0 + q_\varepsilon r \leq \langle \xi_\varepsilon, z \rangle_{(X^*, X)} + p_\varepsilon z^0 - q_\varepsilon L\sqrt{\varepsilon} \]
for each \( (z, z^0, r) \in A_\varepsilon \) and for each \( (\varepsilon, z^0, -L\sqrt{\varepsilon}) \in B_\varepsilon \).

Clearly \( (0, 0, -L\sqrt{\varepsilon}) \in B_\varepsilon \) (this corresponds to the case when no variations are made) and \((0, 0, 1/n) \in A_\varepsilon \) for each positive integer \( n \). The latter implies that
\[ \sup \left\{ \langle \xi_\varepsilon, z \rangle_{(X^*, X)} + p_\varepsilon z^0 + q_\varepsilon r : (z, z^0, r) \in A_\varepsilon \right\} \geq 0. \]

If we assume that there exists \( (z, z^0, r) \in A_\varepsilon \) with
\[ \langle \xi_\varepsilon, z \rangle_{(X^*, X)} + p_\varepsilon z^0 + q_\varepsilon r > 0, \]
it follows that \( nz, nz^0, nr \in A_\varepsilon \) for each positive integer \( n \) and
\[ \langle \xi_\varepsilon, nz \rangle_{(X^*, X)} + p_\varepsilon nz^0 + q_\varepsilon nr \to \infty \text{ as } n \to \infty. \]

The last conclusion contradicts (3.8). Hence
\[ \langle \xi_\varepsilon, z \rangle_{(X^*, X)} + p_\varepsilon z^0 + q_\varepsilon r \leq 0 \leq \langle \xi_\varepsilon, z \rangle_{(X^*, X)} + p_\varepsilon z^0 - q_\varepsilon L\sqrt{\varepsilon} \]
for each \( (z, z^0, r) \in A_\varepsilon \) and for each \( (\varepsilon, z^0, -L\sqrt{\varepsilon}) \in B_\varepsilon \).

Let us fix an arbitrary \( \theta \in \Xi \) and set \( \Gamma_{\varepsilon, \theta} := a_\varepsilon d_S^0(x_\varepsilon(T); z_\varepsilon^0(T)) + b_\varepsilon z_\varepsilon^0. \)
Then we obtain according to (3.9) that
\[ \langle \xi_\varepsilon, z_\varepsilon^0(T) \rangle_{(X^*, X)} + p_\varepsilon z_\varepsilon^0 + q_\varepsilon \Gamma_{\varepsilon, \theta} \leq \langle \xi_\varepsilon, z_\varepsilon^0(T) \rangle_{(X^*, X)} + p_\varepsilon z_\varepsilon^0 - q_\varepsilon L\sqrt{\varepsilon}, \]
i.e.,

\[ q \varepsilon \left( a \varepsilon \, d^0_S \left( y \varepsilon (T); z^0 \varepsilon (T) \right) \right) + b \varepsilon \, z^0,0 \varepsilon \leq -q \varepsilon L \sqrt{\varepsilon}. \]

Since \( \theta \in \Xi \), inequality (3.7) holds true,

\[ a \varepsilon \, d^0_S \left( x \varepsilon (T); z^0 \varepsilon (T) \right) + b \varepsilon \, z^0,0 \varepsilon \geq -L \sqrt{\varepsilon}, \]

and, hence, \( q \varepsilon \leq 0 \). If we assume that \( q \varepsilon = 0 \), then we have that

\[ \langle \xi \varepsilon, z \rangle_{(X^*, X)} + p \varepsilon \, z^0 \leq 0 \]

for each \( z \in X \) and each \( z^0 \in \mathbb{R}^1 \). This implies that \( \xi \varepsilon = 0 \) and \( p \varepsilon = 0 \), which is impossible. Hence \( q \varepsilon < 0 \) and we can think without loss of generality that \( q \varepsilon = -1 \).

Thus we obtain according to (3.9) that

\[ 0 \leq \langle \xi \varepsilon, z \rangle_{(X^*, X)} + p \varepsilon \, z^0 \leq 0 \]

for each \( z \in X \) and each \( z^0 \in \mathbb{R}^1 \). This implies that \( \xi \varepsilon = 0 \) and \( p \varepsilon = 0 \), which is impossible. Hence \( q \varepsilon < 0 \) and we can think without loss of generality that \( q \varepsilon = -1 \).

Thus we obtain according to (3.9) that

\[ 0 \leq \langle \xi \varepsilon, z \rangle_{(X^*, X)} + p \varepsilon \, z^0 \leq 0 \]

for each \( z \in X \) and each \( z^0 \in \mathbb{R}^1 \). This implies that \( \xi \varepsilon = 0 \) and \( p \varepsilon = 0 \), which is impossible. Hence \( q \varepsilon < 0 \) and we can think without loss of generality that \( q \varepsilon = -1 \).

Thus we obtain according to (3.9) that

\[ 0 \leq \langle \xi \varepsilon, z \rangle_{(X^*, X)} + p \varepsilon \, z^0 \leq 0 \]

for each \( z \in X \) and each \( z^0 \in \mathbb{R}^1 \). This implies that \( \xi \varepsilon = 0 \) and \( p \varepsilon = 0 \), which is impossible. Hence \( q \varepsilon < 0 \) and we can think without loss of generality that \( q \varepsilon = -1 \).

Thus we obtain according to (3.9) that

\[ 0 \leq \langle \xi \varepsilon, z \rangle_{(X^*, X)} + p \varepsilon \, z^0 \leq 0 \]

for each \( z \in X \) and each \( z^0 \in \mathbb{R}^1 \). This implies that \( \xi \varepsilon = 0 \) and \( p \varepsilon = 0 \), which is impossible. Hence \( q \varepsilon < 0 \) and we can think without loss of generality that \( q \varepsilon = -1 \).

Thus we obtain according to (3.9) that

\[ 0 \leq \langle \xi \varepsilon, z \rangle_{(X^*, X)} + p \varepsilon \, z^0 \leq 0 \]

for each \( z \in X \) and each \( z^0 \in \mathbb{R}^1 \). This implies that \( \xi \varepsilon = 0 \) and \( p \varepsilon = 0 \), which is impossible. Hence \( q \varepsilon < 0 \) and we can think without loss of generality that \( q \varepsilon = -1 \).

Thus we obtain according to (3.9) that

\[ 0 \leq \langle \xi \varepsilon, z \rangle_{(X^*, X)} + p \varepsilon \, z^0 \leq 0 \]

for each \( z \in X \) and each \( z^0 \in \mathbb{R}^1 \). This implies that \( \xi \varepsilon = 0 \) and \( p \varepsilon = 0 \), which is impossible. Hence \( q \varepsilon < 0 \) and we can think without loss of generality that \( q \varepsilon = -1 \).
Take an arbitrary \( x_1 \). Then in the case \( a_\varepsilon \neq 0 \) we have
\[
d_S(x_1) - d_S(x_\varepsilon(T)) \geq \left< \frac{\xi_\varepsilon}{a_\varepsilon}, x_1 - x_\varepsilon(T) \right>
\]
because \( \frac{\xi_\varepsilon}{a_\varepsilon} \in \partial d_S(x_\varepsilon(T)) \) and \( d_S(\cdot) \) is convex (since the set \( S \) is convex). For \( x_1 \in S \) we obtain
\[
d_S(x_\varepsilon(T)) + \left< \frac{\xi_\varepsilon}{a_\varepsilon}, x_1 - x_\varepsilon(T) \right> \leq 0.
\]
Multiply by \( a_\varepsilon = \frac{d_S(x_\varepsilon(T))}{J_\varepsilon(u_\varepsilon(\cdot), v_\varepsilon)} \). Then
\[
\tag{3.11}
\frac{d_S^2(x_\varepsilon(T))}{J_\varepsilon(u_\varepsilon(\cdot), v_\varepsilon)} + \left< \xi_\varepsilon, x_1 - x_\varepsilon(T) \right> \leq 0.
\]
The above written inequality also remains true in the case when \( a_\varepsilon = 0 \).

**Step 7. The limit case as \( \varepsilon \) tends to zero.** This is a technical step aiming at the limit case as \( \varepsilon \) tends to zero of the separation inequality obtained in the previous step.

From Ekeland’s variational principle we have that
\[
\text{(E1)} \quad \tilde{d}(\{u_\varepsilon(\cdot), v_\varepsilon\}, (\bar{u}(\cdot), \bar{v})) \leq \sqrt{\varepsilon},
\]
\[
\text{(E2)} \quad J_\varepsilon(\{u_\varepsilon(\cdot), v_\varepsilon\}, (u(\cdot), v)) \geq J_\varepsilon(\{u_\varepsilon(\cdot), v_\varepsilon\})
\]
for every \((u(\cdot), v) \in \mathcal{U} \times \mathcal{V}\). For every \( \theta = (\{u_i(\cdot)\}_{i=1}^k, \{v_i\}_{i=1}^k, \{\lambda_i\}_{i=1}^k) \in \Xi \) we define
\[
\begin{align*}
\tag{3.12}
z_\theta(t) := & \varphi'(\bar{v}) \circ \left( \sum_{i=1}^k \lambda_i v_i - \bar{v} \right) + \int_0^t f_x'(s, \bar{x}(s), \bar{u}(s), \bar{v}) \, z_\theta(s) \, ds \\
& + \sum_{i=1}^k \lambda_i \int_0^t \left( f(s, \bar{x}(s), u_i(s), \bar{v}) - f(s, \bar{x}(s), \bar{u}(s), \bar{v}) \right) \, ds \\
& + \int_0^t f_v'(s, \bar{x}(s), \bar{u}(s), \bar{v}) \circ \left( \sum_{i=1}^k \lambda_i v_i - \bar{v} \right) \, ds,
\end{align*}
\]
\[
\begin{align*}
\tag{3.13}
z_\theta^0(t) := & \int_0^T f_x^{0\theta}(s, \bar{x}(s), \bar{u}(s)) \, z_\theta(s) \, ds \\
& + \sum_{i=1}^k \lambda_i \int_0^T \left( f_0^{0}(s, \bar{x}(s), u_i(s), \bar{v}) - f_0^{0}(s, \bar{y}(s), \bar{u}(s), \bar{v}) \right) \, ds \\
& + \int_0^T f_v^{0\theta}(s, \bar{x}(s), \bar{u}(s), \bar{v}) \circ \left( \sum_{i=1}^k \lambda_i v_i - \bar{v} \right) \, ds.
\end{align*}
\]
Because of (E1) we have that \( \|v_\varepsilon - \bar{v}\| \to 0 \) as \( \varepsilon \to 0 \). This implies that \( \|\varphi'(v_\varepsilon) - \varphi'(v_\varepsilon)\| \to 0 \) as \( \varepsilon \to 0 \).

Also, as in the proof of Lemma 3.4, for a fixed \( t \in [0, T] \) we have
\[
\|z_\varepsilon^0(t) - z_\theta(t)\|_X \leq \left\| (\varphi'(\bar{v}) - \varphi'(v_\varepsilon)) \left( \sum_{i=1}^k \lambda_i v_i - \bar{v} \right) \right\|_X + \left\| \int_0^t f_x'(s, x_\varepsilon(s), u_\varepsilon(s), v_\varepsilon) \left( z_\theta^0(s) - z_\theta(s) \right) \, ds \right\|_X
\]
Applying the Gronwall inequality, we obtain
\[
\|z_{\varepsilon}^\theta(t) - z_\theta(t)\|_X \leq e^{MT} a_{\varepsilon}(1)_{\epsilon \to 0}.
\]
Thus the above convergence is uniform in \( t \in [0, T] \) and \( \theta \in \Xi \). One can repeat the same calculations to obtain that \( \|z_{\varepsilon}^{\theta,0} - z_\theta^{\theta,0}\|_{\Xi} \to 0 \) uniformly in \( \theta \in \Xi \).

From (3.11) we obtain \( \langle \xi, x_1 - x_\varepsilon(T) \rangle \leq 0 \) for each point \( x_1 \) from \( S \). Therefore,
\[
(3.14) \quad \langle \xi, x_1 - x(T, \bar{u}(\cdot), \bar{v}) \rangle = \langle \xi, x_1 - x_\varepsilon(T) \rangle + \langle \xi, x_\varepsilon(T) - x(T, \bar{u}(\cdot), \bar{v}) \rangle
\]
\[
\|\xi\|_X \cdot \|x_\varepsilon(T) - x(T, \bar{u}(\cdot), \bar{v})\|_X \leq \|x_\varepsilon(T) - x(T, \bar{u}(\cdot), \bar{v})\|_X =: \phi_\varepsilon \to 0.
\]

Using (3.10), we obtain
\[
\langle \xi, z_\theta(T) \rangle + p_\varepsilon z_\theta^{\theta,0} = \langle \xi, z_\varepsilon^\theta(T) \rangle + p_\varepsilon z_\varepsilon^{\theta,0} - \langle \xi, z_\varepsilon^\theta(T) - z_\theta(T) \rangle + p_\varepsilon (z_\theta^{\theta,0} - z_\varepsilon^{\theta,0})
\]
\[
\geq -L\sqrt{\varepsilon} - \|\xi\|_X \cdot \|z_\varepsilon^\theta(T) - z_\theta(T)\|_X - |p_\varepsilon| \|z_\theta^{\theta,0} - z_\varepsilon^{\theta,0}\|
\]
\[
\geq -L\sqrt{\varepsilon} - \|z_\varepsilon^\theta(T) - z_\theta(T)\|_X - |z_\theta^{\theta,0} - z_\varepsilon^{\theta,0}|\]
(because \( \|\xi\|_X \leq 1 \) and \( |p_\varepsilon| \leq 1 \)). From the last inequality we subtract (3.14) and obtain
\[
\langle \xi, z_\theta(T) - (x_1 - x(T, \bar{u}(\cdot), \bar{v})) \rangle + p_\varepsilon z_\theta^{\theta,0} \geq -L\sqrt{\varepsilon} - \phi_\varepsilon - \|z_\varepsilon^\theta(T) - z_\theta(T)\|_X - |z_\theta^{\theta,0} - z_\varepsilon^{\theta,0}|.
\]
Thus
\[
(3.15) \quad \langle \xi, z_\theta(T) - (x_1 - x(T, \bar{u}(\cdot), \bar{v})) \rangle + p_\varepsilon z_\theta^{\theta,0} \geq -\kappa_\varepsilon
\]
with
\[
\kappa_\varepsilon \geq L\sqrt{\varepsilon} + \phi_\varepsilon + \|z_\varepsilon,\theta(T) + z_\theta(T)\|_X + |z_\theta^{\theta,0} - z_\varepsilon^{\theta,0}|\] and \( \kappa_\varepsilon \epsilon \to 0 \),
and the convergence is uniform with respect to \( \theta \in \Xi \).
Let $G(\cdot, \cdot)$ be the evolution operator generated by $f_x^\prime(\cdot, \bar{x}(\cdot), \bar{u}(\cdot), \bar{v})$. Then (3.12) implies

$$z_\theta(t) = G(t, 0) \varphi'(\bar{v}) \left( \sum_{i=1}^{k} \lambda_i v_i - \bar{v} \right)$$

(3.16) 

$$+ \int_{0}^{t} G(t, s) \left( \sum_{i=1}^{k} \lambda_i \left( f(s, \bar{x}(s), u_i(s)) - f(s, \bar{x}(s), \bar{u}(s)) \right) \right) \, ds$$

$$+ \int_{0}^{t} G(t, s) \left( f'_v(s, \bar{x}(s), \bar{u}(s), \bar{v}) \left( \sum_{i=1}^{k} \lambda_i v_i - \bar{v} \right) \right) \, ds.$$ 

Clearly, $P = \{ z_\theta(\cdot) : \theta \in \Xi \}$. Hence 

$$P - S + \bar{x}(T) = \{ \zeta : \zeta = z_\theta(T) - (x_1 - \bar{x}(T)) \text{ for some } \theta \in \Xi \text{ and } x_1 \in S \}.$$ 

Therefore, (3.15) implies

$$\langle \xi_\epsilon, \zeta \rangle + p_\epsilon z^0_\theta \geq -\kappa_\epsilon$$

for each $\zeta \in P - S + \bar{y}(T)$. 

Up to now we have proved that for each $\epsilon > 0$ there exist $\xi_\epsilon \in X^*$, $p_\epsilon \in \mathbb{R}^1$, and $\kappa_\epsilon > 0$ such that $\kappa_\epsilon \rightarrow 0$, $\|\xi_\epsilon\|^2 + p_\epsilon^2 = 1$, and

(3.17) 

$$\langle \xi_\epsilon, z_\theta(T) - (x_1 - \bar{x}(T)) \rangle + p_\epsilon z^0_\theta \geq -\kappa_\epsilon$$

for every $x_1 \in S$ and for each $\theta \in \Xi$. 

Let us consider the sequence $\{\xi_{1/n}\}_{n=1}^\infty$, and let $\{\xi_\delta\}_{\delta \in \Delta}$ be a $w^*$-convergent submesh. Let $\xi_\delta \rightarrow \xi \in X^*$. Without loss of generality we may think that the mesh $\{p_\delta\}_{\delta \in \Delta}$ is convergent. Let $p_\delta \rightarrow \bar{p} \in \mathbb{R}^1$. If $\bar{p} \neq 0$, then $\langle \xi, \bar{p} \rangle \neq (0, 0)$. If $\bar{p} = 0$, then $p_\delta \rightarrow 0$ and so

$$\langle \xi_\delta, z_\theta(T) - (x_1 - \bar{x}(T)) \rangle \geq -p_\delta z^0_\theta - \kappa_\delta \rightarrow 0$$

uniformly with respect to $\theta \in \Xi$ and $x_1 \in S$. Two cases are possible: If the closed affine hull of $P - S$ does not coincide with $X$, then there exists a norm one functional which we denote again by $\xi$ which vanishes on $P - S + \bar{x}(T)$ (according to the Hahn–Banach theorem). Thus we have

$$\langle \bar{\xi}, z_\theta(T) - (x_1 - \bar{x}(T)) \rangle = 0$$

for every $x_1 \in S$ and for each $\theta \in \Xi$. 

Otherwise, the linear span of the set $P - S$ is dense in $X$, and, therefore, it is finite-codimensional in $X$ according to Definition 1.5 in [22]. Then applying Lemma 3.6 from [22] (because $\|\xi_\delta\|^2 = 1 - p^2_\delta > 1/2$ for all sufficiently large $\delta$), we obtain that $\xi \neq 0$. Thus in all cases we have that $\langle \bar{\xi}, \bar{p} \rangle \neq (0, 0)$. Also, taking a limit in (3.17), we obtain that

(3.18) 

$$\langle \bar{\xi}, z_\theta(T) - (x_1 - \bar{x}(T)) \rangle + \bar{p} z^0_\theta \geq 0$$

for every $x_1 \in S$ and for each $\theta \in \Xi$. 

A PONTRYAGIN MAXIMUM PRINCIPLE
Step 8. Deriving the maximum principle in integral form. Here we introduce the adjoint variable and derive the integral maximum property for the Hamiltonian and a transversality condition.

We denote by $\psi$ the unique solution of

$$\psi(t) = \psi(T) + \int_t^T (f_x' (s, \bar{x}(s), \bar{u}(s), \bar{v}))^* \psi(s) \, ds + \int_t^T \psi^0 f^0_x (s, \bar{x}(s), \bar{u}(s), \bar{v}) \, ds,$$

where $\psi^0 := -\bar{p}$ and $\psi(T) := -\bar{\xi}$. According to Proposition 5.7 [22, Chap. 2],

$$\psi(t) = G^* (T, t) \psi(T) + \psi^0 \int_t^T G^* (s, t) f^0_x (s, \bar{x}(s), \bar{u}(s), \bar{v}) \, ds.$$

We write (3.18) as

$$(3.19) \quad \langle -\psi(T), x_1 - \bar{x}(T) - z_\theta(T) \rangle + \psi^0 z_\theta^0 \leq 0$$

for every $x_1 \in S$ and for each $\theta \in \Xi$. Using (3.16), we obtain

$$\langle \psi(T), z_\theta(T) \rangle - \langle \psi(0), z_\theta(0) \rangle$$

$$= \int_0^T \frac{d}{dt} \langle \psi(t), z_\theta(t) \rangle \, dt = \int_0^T \left( \left\langle \dot{\psi}(t), z_\theta(t) \right\rangle + \langle \psi(t), \dot{z}_\theta(t) \rangle \right) \, dt$$

$$= \int_0^T \left\langle - (f_x' (t, \bar{x}(t), \bar{u}(t), \bar{v}))^* \psi(t) - \psi^0 f^0_x (t, \bar{x}(t), \bar{u}(t), \bar{v}), \ z_\theta(t) \right\rangle \, dt$$

$$+ \int_0^T \left\langle \psi(t), f_x' (t, \bar{x}(t), \bar{u}(t), \bar{v}) z_\theta(t) + \sum_{i=1}^k \lambda_i \left( f(t, \bar{x}(t), u_i(t), \bar{v}) - f(t, \bar{x}(t), \bar{u}(t), \bar{v}) \right) \right. \left. + f_x' (t, \bar{x}(t), \bar{u}(t), \bar{v}) \circ \left( \sum_{i=1}^k \lambda_i (v_i - \bar{v}) \right) \right\rangle \, dt$$

$$= - \int_0^T \langle \psi(t), f_x' (t, \bar{x}(t), \bar{u}(t), \bar{v}) z_\theta(t) \rangle \, dt - \psi^0 \int_0^T \langle f^0_x (t, \bar{x}(t), \bar{u}(t), \bar{v}), z_\theta(t) \rangle \, dt$$

$$+ \int_0^T \langle \psi(t), f_x' (t, \bar{x}(t), \bar{u}(t), \bar{v}) z_\theta(t) \rangle \, dt$$

$$+ \int_0^T \left\langle \psi(t), f_x' (t, \bar{x}(t), \bar{u}(t), \bar{v}) \circ \left( \sum_{i=1}^k \lambda_i (v_i - \bar{v}) \right) \right\rangle \, dt$$

$$+ \sum_{i=1}^k \lambda_i \int_0^T \langle \psi(t), f(t, \bar{x}(t), u_i(t), \bar{v}) - f(t, \bar{x}(t), \bar{u}(t), \bar{v}) \rangle \, dt.$$
Hence (because of (3.13))

\[
\langle \psi(T), z_\theta(T) \rangle - \langle \psi(0), z_\theta(0) \rangle + \psi^0 z^0_\theta
\]

\[
= -\psi^0 \int_0^T \langle f^0_\nu(t, \bar{x}(t), \bar{u}(t), v), z_\theta(t) \rangle \, dt
\]

\[
+ \sum_{i=1}^k \lambda_i \int_0^T \langle \psi(t), f(t, \bar{x}(t), u_i(t), \bar{v}) - f(t, \bar{x}(t), \bar{u}(t), \bar{v}) \rangle \, dt
\]

\[
+ \int_0^T \langle \psi(t), f^0_\nu(t, \bar{x}(t), \bar{u}(t), \bar{v}) \rangle \, dt
\]

\[
+ \psi^0 \sum_{i=1}^k \lambda_i \int_0^T \langle f^0(t, \bar{x}(t), u_i(t), \bar{v}) - f^0(t, \bar{x}(t), \bar{u}(t), \bar{v}) \rangle \, dt
\]

\[
= \int_0^T \left( \langle \psi(t), \sum_{i=1}^k \lambda_i f(t, \bar{x}(t), u_i(t), \bar{v}) \rangle + \psi^0 \sum_{i=1}^k \lambda_i f^0(t, \bar{x}(t), u_i(t), \bar{v}) \right)
\]

\[
+ \sum_{i=1}^k \lambda_i \left( \langle \psi(t), f^0_\nu(t, \bar{x}(t), \bar{u}(t), \bar{v}) \rangle + \psi^0 f^0_\nu(t, \bar{x}(t), \bar{u}(t), \bar{v}) \rangle v_i \right) \, dt
\]

\[- \int_0^T \langle \psi(t), f(t, \bar{x}(t), \bar{u}(t), \bar{v}) \rangle + \psi^0 f^0(t, \bar{x}(t), \bar{u}(t), \bar{v}) \rangle \, dt
\]

\[+ \langle \psi(t), f^0_\nu(t, \bar{x}(t), \bar{u}(t), \bar{v}) \rangle + \psi^0 f^0_\nu(t, \bar{x}(t), \bar{u}(t), \bar{v}) \rangle \, dt.\]

Let us define

\[H(t, x, u, v, \psi^0, \psi) = \langle \psi, f(t, x, u, v) \rangle + \psi^0 f^0(t, x, u, v).\]

Take \( k = 1, v_1 := v = \bar{v}, \) and \( u(\cdot) \equiv \bar{u}. \) Then

\[\langle \psi(T), x_1 - \bar{x}(T) \rangle \geq 0\]

for each \( x_1 \in S. \) This relation is called the transversality condition.

Take \( k = 1, v_1 := v = \bar{v}, \) and \( u(\cdot) \) to be an arbitrary element of \( U. \) Then

\[\langle \psi(T), z_\theta(T) \rangle - \langle \psi(0), z_\theta(0) \rangle + \psi^0 z^0_\theta\]

\[= \int_0^T \left( H(t, \bar{x}(t), u(t), \bar{v}, \psi^0, \psi(t)) - H(t, \bar{x}(t), \bar{u}(t), \bar{v}, \psi^0, \psi(t)) \right) \, dt.\]

On the other hand, if we substitute \( x_1 \) by \( \bar{x}(T) \) in the inequality (3.19),

\[\langle -\psi(T), x_1 - \bar{x}(T) - z_\theta(T) \rangle + \psi^0 z^0_\theta \leq 0,\]
then we obtain
\[ \langle \psi(T), z_\theta(T) \rangle + \psi^0 z_\theta^0 \leq 0 \]
for each \( \theta \in \Xi \). Since \( z_\theta(0) = \varphi'(\bar{v})(v - \bar{v}) \), with \( v = \bar{v} \), we have
\[ 0 \geq \int_0^T \left( H(t, \bar{x}(t), u(t), \bar{v}, \psi^0, \psi(t)) - H(t, \bar{x}(t), \bar{u}(t), \bar{v}, \psi^0, \psi(t)) \right) dt \]
for every \( u(\cdot) \in \mathcal{U} \).

**Step 9. Deriving the maximum principle.** Here we derive the pointwise maximum property for the Hamiltonian.

Since \( U \) is separable there exists a countable and dense set
\[ U_0 = \{ u_i \}_{i=1}^\infty \subset U. \]
For each \( u_i \in U_0 \) we set
\[ g_i(s) = H(s, \bar{x}(s), u_i, \bar{v}, \psi^0, \psi(s)) - H(s, \bar{x}(s), \bar{u}(s), \bar{v}, \psi^0, \psi(s)). \]
Then \( g_i(\cdot) \in L^1(0, T) \). Thus there exists a measurable set \( F_i \subset [0, T] \) with meas \( F_i = T \) such that every point in \( F_i \) is a Lebesgue point of \( g_i \), namely,
\[ \lim_{\mu \to 0} \frac{1}{2\mu} \int_{t-\mu}^{t+\mu} (g_i(s) - g_i(t)) ds = 0 \]
for each \( t \in F_i \). Now, for any \( t \in F_i \) and for each \( \mu > 0 \) we define
\[ u_\mu(s) := \begin{cases} \bar{u}(s) & \text{if } |s - t| > \mu, \\ u_i & \text{if } |s - t| \leq \mu. \end{cases} \]
Then from (3.20) we obtain for each \( \mu > 0 \) that (by replacing \( u \) by \( u_\mu \))
\[ \int_{t-\mu}^{t+\mu} g_i(s) ds \leq 0 \]
because \( u_\mu \equiv \bar{u} \) on the set \([0, T] \setminus (t - \mu, t + \mu)\). Since
\[ g_i(t) = \lim_{\mu \to 0} \frac{1}{2\mu} \int_{t-\mu}^{t+\mu} g_i(s) ds \leq 0, \]
we have that
\[ H(t, \bar{x}(t), u_i, \bar{v}, \psi^0, \psi(t)) \leq H(t, \bar{x}(t), \bar{u}(t), \bar{v}, \psi^0, \psi(t)) \]
for any \( t \in F = \bigcap_{j=1}^\infty F_j \) (meas \( F = T \)), and for each \( u_i \in U_0 \). By the continuity of the Hamiltonian in \( u \) we obtain that
\[ H(t, \bar{x}(t), \bar{u}(t), \bar{v}, \psi^0, \psi(t)) = \max_{u \in \mathcal{U}} H(t, \bar{x}(t), u, \bar{v}, \psi^0, \psi(t)) \text{ a.e. in } [0, T]. \]

Take again \( k = 1 \) and \( u(\cdot) = \bar{u}(\cdot) \). If we substitute \( x_1 \) by \( \bar{x}(T) \) in the inequality (3.19), we obtain
\[ \langle \psi(T), z_\theta(T) \rangle + \psi^0 z_\theta^0 \leq 0 \]
for each \( \theta \in \Xi \). Also, we substitute \( z_0^0 \) by

\[
\int_0^T f^0_x(t, \bar{x}(t), \bar{u}(t), \bar{v}) \left[ G(t, 0) \varphi'(\bar{v}) + \int_0^t G(t, s) f'_x(s, \bar{x}(s), \bar{u}(s), \bar{v}) \, ds \right] \circ (v - \bar{v}) \, dt
\]

and obtain

\[
z_0(T) \text{ by } G(T, 0) \varphi'(\bar{v}) (v - \bar{v}) + \int_0^T G(T, s) f'_x(s, \bar{x}(s), \bar{u}(s), \bar{v}) (v - \bar{v}) \, ds
\]

and obtain

\[
\begin{aligned}
\left\langle \psi(T), G(T, 0) \varphi'(\bar{v}) (v - \bar{v}) + \int_0^T G(T, s) f'_x(s, \bar{x}(s), \bar{u}(s), \bar{v}) (v - \bar{v}) \, ds \rightangle \\
+ \psi_0 \int_0^T f^0_x(t, \bar{x}(t), \bar{u}(t), \bar{v}) \left[ G(t, 0) \varphi'(\bar{v}) + \int_0^t G(t, s) f'_x(s, \bar{x}(s), \bar{u}(s), \bar{v}) \, ds \right] \circ (v - \bar{v}) \, dt
\end{aligned}
\]

+ \psi_0 \int_0^T f^0_x(t, \bar{x}(t), \bar{u}(t), \bar{v}) (v - \bar{v}) \, dt \leq 0

for each \( v \in \mathcal{V} \). Changing the order of integration, we next obtain

\[
\begin{aligned}
\left\langle G^*(T, 0) \psi(T) + \psi_0 \int_0^T G^*(t, 0) f^0_x(t, \bar{x}(t), \bar{u}(t), \bar{v}) \, dt, \varphi'(\bar{v}) (v - \bar{v}) \rightangle \\
+ \int_0^T \left\langle G^*(T, t) \psi(T), f'_x(t, \bar{x}(t), \bar{u}(t), \bar{v}) (v - \bar{v}) \right \rangle \, dt + \psi_0 \int_0^T f^0_x(t, \bar{x}(t), \bar{u}(t), \bar{v}) (v - \bar{v}) \, dt
\end{aligned}
\]

+ \psi_0 \int_0^T \left\langle \int_t^T G^*(s, t) f^0_x(s, \bar{x}(s), \bar{u}(s), \bar{v}) \, ds, f'_x(t, \bar{x}(t), \bar{u}(t), \bar{v}) (v - \bar{v}) \right \rangle \, dt \leq 0

for each \( v \in \mathcal{V} \). Since

\[\psi(t) = G^*(T, t) \psi(T) + \psi_0 \int_t^T G^*(s, t) f^0_x(s, \bar{x}(s), \bar{u}(s), \bar{v}) \, ds,\]

we obtain (applying it for \( t = 0 \) as well as for arbitrary \( t \in [0, T] \))

\[
\left\langle \psi(0), \varphi'(\bar{v}) (v - \bar{v}) \right\rangle + \int_0^T \left\langle \psi(t), f'_x(t, \bar{x}(t), \bar{u}(t), \bar{v}) (v - \bar{v}) \right \rangle \, dt
\]

+ \psi_0 \int_0^T f^0_x(t, \bar{x}(t), \bar{u}(t), \bar{v}) (v - \bar{v}) \, dt \leq 0

for each \( v \in \mathcal{V} \). Finally, we obtain

\[
\left\langle \varphi'(\bar{v}) \psi(0) + \int_0^T H_v'(t, \bar{x}(t), \bar{u}(t), \bar{v}) \, dt, (v - \bar{v}) \right\rangle \leq 0
\]
follows as in the proof of the theorem. For some specific right-hand sides of the state equation and integrands of the cost functional this assumption can be omitted. For example, if

$$f^0(t, x, u, v) = \sum_{i=1}^{N_0} f_i^0(t, x, v) g_i^0(u) + f_{N_0+1}^0(t, x, v),$$

$$f(t, x, u, v) = \sum_{i=1}^{N} f_i(t, x, v) g_i(u) + f_{N+1}(t, x, v),$$

then the conclusion of the main theorem holds true without the separability assumption on $U$. Indeed, the inequality (3.20) can be written as follows

$$0 \geq \int_0^T \left\langle \psi(t), \sum_{i=1}^{N} \left( f_i(t, \bar{x}(t), \bar{v}) g_i(u(t)) - f_i(t, \bar{x}(t), \bar{v}) g_i(\bar{u}(t)) \right) \right\rangle dt \geq\right.$$

$$(3.21) + \psi_0 \int_0^T \sum_{i=1}^{N_0} \left( f_i^0(t, \bar{x}(t), \bar{v}) g_i^0(u(t)) - f_i^0(t, \bar{x}(t), \bar{v}) g_i^0(\bar{u}(t)) \right) dt.$$

For $(s, u) \in [0, T] \times U$ we define

$$h(s, u) := \left\langle \psi(s), \sum_{i=1}^{N} \left( f_i(s, \bar{x}(s), \bar{v}) g_i(u) - f_i(s, \bar{x}(s), \bar{v}) g_i(\bar{u}(s)) \right) \right\rangle + \psi_0 \sum_{i=1}^{N_0} \left( f_i^0(s, \bar{x}(s), \bar{v}) g_i^0(u) - f_i^0(s, \bar{x}(s), \bar{v}) g_i^0(\bar{u}(s)) \right).$$

Since

$$\lim_{\mu \to 0} \frac{1}{2\mu} \int_{t-\mu}^{t+\mu} h(s, u) \, ds = \sum_{i=1}^{N} \left( \lim_{\mu \to 0} \frac{1}{2\mu} \int_{t-\mu}^{t+\mu} \langle \psi(s), f_i(s, \bar{x}(s), \bar{v}) \rangle \, ds \right) g_i(u) - \lim_{\mu \to 0} \frac{1}{2\mu} \int_{t-\mu}^{t+\mu} \langle \psi(s), \sum_{i=1}^{N} f_i(s, \bar{x}(s), \bar{v}) g_i(\bar{u}(s)) \rangle \, ds$$

$$+ \psi_0 \sum_{i=1}^{N_0} \left( \lim_{\mu \to 0} \frac{1}{2\mu} \int_{t-\mu}^{t+\mu} f_i^0(s, \bar{x}(s), \bar{v}) \, ds \right) g_i^0(u)$$

$$- \psi_0 \lim_{\mu \to 0} \frac{1}{2\mu} \int_{t-\mu}^{t+\mu} \left( \sum_{i=1}^{N_0} f_i^0(s, \bar{x}(s), \bar{v}) g_i^0(\bar{u}(s)) \, ds \right),$$

the Lebesgue points of $h(\cdot, u)$ do not depend on the choice of the control $u$. Because $h(\cdot, u) \in L^1(0, T)$, there exists a measurable set $F \subset [0, T]$ with meas $F = T$ such that every point in $F$ is a Lebesgue point of $h(\cdot, u)$ for every $u \in U$. Then the conclusion follows as in the proof of the theorem.
4. An example: Optimal age-specific education. The following example is a particular case of the model developed and analyzed in [26]:

\[
\max \int_0^T e^{-rt} \int_0^\omega \left[ \theta_L(t) \pi_L(s) - \theta_H(t) \pi_H(s) - p(w(t,s)) \right] L(t,s) \, ds \, dt
\]

subject to

\[
L_t + L_s = (-\delta(t,s) - w(t,s))L(t,s) + \delta(t,s)N(t,s) \text{ a.e. in } [0,T] \times [0,\omega],
\]

\[
L(t,0) \geq \alpha \quad \text{on } [0,T] \text{ – given boundary data,}
\]

\[
L(0,s) \geq \alpha \quad \text{on } [0,\omega] \text{ – given initial data,}
\]

\[
w(t,s) \geq 0 \quad \text{a.e. in } [0,T] \times [0,\omega].
\]

Here \( \alpha > 0, r > 0, \delta(\cdot), \theta_L(\cdot), \theta_H(\cdot), \pi_L(\cdot), \pi_H(\cdot), \text{and } p(\cdot) \) are data, \( N(\cdot) \) turns out to be data, too, as will be explained (all functional data are assumed continuous and with positive values), \( L(\cdot, \cdot) \) is the state variable, and \( w(\cdot, \cdot) \) is the control. In this example \( t \) means time, \( s \) means age, and the domain where they vary is the rectangle \([0,T] \times [0,\omega]\). In the age-structured models, of which the above example is a representative, the end of the time horizon \( T \) is typically assumed to be much larger than the upper bound \( \omega \) of the age. We also mention that there are two state variables, \( L(\cdot, \cdot) \) and \( H(\cdot, \cdot) \), in the model considered in [26]. However, as is pointed out there, the dynamics is such that \( L(t,s) + H(t,s) = N(t,s) \) on \([0,T] \times [0,\omega]\), where \( N(\cdot) \) is completely determined by the initial and boundary data, and so it can be considered data itself. Therefore, the state variables can be reduced to one—either \( L(\cdot, \cdot) \) or \( H(\cdot, \cdot) \); here we have chosen \( L(\cdot, \cdot) \). We refer the reader to [26] for more details.

We add a (pointwise) terminal state constraint to (4.1):

\[
L(t,\omega) = g(t) \text{ for each } t \in [\omega, T],
\]

where \( g(\cdot) \in L_\infty[\omega, T] \) is given. In terms of the model (cf. [26]), since \( L(t,s) \) denotes the number of people of age \( s \) at time \( t \) in a specific group, adding (4.2) to the considered problem means that we add the requirement that the number of people \( L(t,\omega) \) leaving the group (due to retirement) has a given time-profile \( g(t) \).

In order to embed this problem into the abstract framework of the previous sections, we first change the independent variables: in the subdomain \( D_1 := \{(t,s) : 0 \leq t \leq \omega, t \leq s \leq \omega\} \) (Figure 1) we transform the coordinates into

\[
\begin{align*}
\tau & = t, \\
\sigma & = t - s,
\end{align*}
\]

and in the subdomain \( D_2 := \{(t,s) : 0 \leq t \leq T, 0 \leq s \leq \min\{t,\omega\}\} \) (Figure 1) we transform them into

\[
\begin{align*}
\tau & = s, \\
\sigma & = t - s.
\end{align*}
\]

We thus obtain the domain of the new variables to be the area \( \Omega := \{(\tau,\sigma) : \tau \in [0,\omega], \sigma \in [\tau - \omega, T - \tau]\} \) being the union of the image \( \Omega_1 := \{(\tau,\sigma) : \tau \in [0,\omega], \sigma \in [\tau - \omega, 0]\} \) of \( D_1 \) and the image \( \Omega_2 := \{(\tau,\sigma) : \tau \in [0,\omega], \sigma \in [0, T - \tau]\} \) of \( D_2 \) (Figure 2).
Let

\[
\begin{align*}
x(\tau, \sigma) := & \begin{cases} 
L(\tau, \tau - \sigma) & \text{if } (\tau, \sigma) \in \Omega_1, \\
L(\tau + \sigma, \tau) & \text{if } (\tau, \sigma) \in \Omega_2,
\end{cases} \\
u(\tau, \sigma) := & \begin{cases} 
\psi(\tau, \tau - \sigma) & \text{if } (\tau, \sigma) \in \Omega_1, \\
\psi(\tau + \sigma, \tau) & \text{if } (\tau, \sigma) \in \Omega_2,
\end{cases}
\end{align*}
\]

be the state and the control variables, respectively, in the new coordinates. The corresponding dynamics is

\[
\dot{x}(\tau, \sigma) := \frac{\partial}{\partial \tau} x(\tau, \sigma) = \hat{f}(\tau, \sigma, x(\tau, \sigma), u(\tau, \sigma)),
\]

where

\[
\hat{f}(\tau, \sigma, x, u) := \begin{cases} 
(-\delta(\tau, \tau - \sigma) - u)x + \delta(\tau, \tau - \sigma)N(\tau, \tau - \sigma) & \text{if } (\tau, \sigma, x, u) \in \Omega_1 \times \mathbb{R}^2, \\
(-\delta(\tau + \sigma, \tau) - u)x + \delta(\tau + \sigma, \tau)N(\tau + \sigma, \tau) & \text{if } (\tau, \sigma, x, u) \in \Omega_2 \times \mathbb{R}^2.
\end{cases}
\]
We next enlarge the area of \((\tau, \sigma)\) to \([0, \omega] \times [-\omega, T]\) and extend the dynamics as follows:

\[
(4.5) \quad f(\tau, \sigma, x, u) := \begin{cases} 
\dot{f}(\tau, \sigma, x, u) & \text{if } (\tau, \sigma, x, u) \in (\Omega_1 \cup \Omega_2) \times R^2, \\
\dot{f}(\sigma + \omega, \sigma, x, u) & \text{if } (\tau, \sigma, x, u) \in \Delta_1 \times R^2, \\
\dot{f}(T - \sigma, \sigma, x, u) & \text{if } (\tau, \sigma, x, u) \in \Delta_2 \times R^2,
\end{cases}
\]

where \(\Delta_1 := \{(\tau, \sigma) : \tau \in [0, \omega], \sigma \in [-\omega, \tau - \omega]\}\) and \(\Delta_2 := \{(\tau, \sigma) : \tau \in [0, \omega], \sigma \in [T - \tau, T]\}\) (Figure 2). We next define \(f^0(\tau, \sigma, x, u) := \)

\[
\begin{align*}
&\left\{ e^{-r\tau} [\theta_L(\tau)p_L(\tau - \sigma) - \theta_H(\tau)p_H(\tau - \sigma) - p(u)] x \right. \text{if } (\tau, \sigma, x, u) \in \Omega_1 \times R^2, \\
&\left. e^{-r(\tau + \sigma)} [\theta_L(\tau + \sigma)p_L(\tau) - \theta_H(\tau + \sigma)p_H(\tau) - p(u)] x \right. \text{if } (\tau, \sigma, x, u) \in \Omega_2 \times R^2,
\end{align*}
\]

and the target set (cf. (4.2)) of the enlarged problem becomes

\[
S = \{ h(\cdot) \in L_\infty[-\omega, T] : h(\sigma) = g(\sigma + \omega) \text{ for almost all } \sigma \in [0, T - \omega] \}.
\]

Then the problem reduces to

\[
\int_0^\omega \int_{\tau - \omega}^{T - \tau} f^0(\tau, \sigma, x(\tau, \sigma), u(\tau, \sigma))d\sigma d\tau \rightarrow \min
\]

subject to

\[
\begin{align*}
\dot{x}(\tau, \sigma) &= f(\tau, \sigma, x(\tau, \sigma), u(\tau, \sigma)) \text{ a.e. in } [0, \omega] \times [-\omega, T], \\
x(0, \sigma) &= x_0(\sigma) \text{ on } [-\omega, T], \\
x(T, \cdot) &\in S, \\
u(\tau, \sigma) &\geq 0 \text{ a.e. in } [0, \omega] \times [-\omega, T].
\end{align*}
\]

The Banach space \(X\) of the space variable in the infinite-dimensional formulation is \(L_\infty[-\omega, T]\); i.e., the new state variable is \(x(\tau) := x(\tau, \cdot), \tau \in [0, \omega]\), with the initial condition \(x_0 = x_0(\cdot)\). Let \(u(\tau) := u(\tau, \cdot), \tau \in [0, \omega]\), be the new control variable. We take the controls to be the Bochner integrable functions of \(\tau \in [0, \omega]\) with values in \(U := L_\infty[-\omega, T]\). Next we set \(f(\tau, \chi, \eta) := f(\tau, \cdot, \chi(\cdot), \eta(\cdot))\) and

\[
f^0(\tau, \chi, \eta) := \int_{\tau - \omega}^{T - \tau} f^0(\tau, \sigma, \chi(\sigma), \eta(\sigma))d\sigma
\]

for \(\tau \in [0, \omega], \chi \in L_\infty[-\omega, T]\), and \(\eta \in U\).

In this way this control problem is formulated as a problem of the type (2.1)–(2.2). We note that \(L_\infty([0, \omega] \times L_\infty[-\omega, T])\) is a proper subset of \(L_\infty([0, \omega] \times [-\omega, T])\). That is why in (4.5) we have extended the dynamics in a relatively complex way instead of simply putting zero outside \((\Omega_1 \cup \Omega_2) \times R^2\), thus ensuring the measurability of \(f\) with respect to \(\tau\). For the same reason the change of the independent variables was chosen as in (4.3) and in (4.4).

The hypothesis (H2) can be checked directly. Regarding the hypothesis (H1), only the separability of the set \(U\) may be not fulfilled, but even without it the conclusion of the main result remains true for our system because of Remark 3.6. Let us verify the
hypothesis \((H_3)\), i.e., that the set \(\mathcal{P} - S\) for this specific control problem is quasi-solid. Let \((\bar{x}(\cdot), \bar{u}(\cdot))\) be the optimal pair.

The linearized equation is

\[
\begin{align*}
\dot{\zeta}(\tau) &= A(\tau)\zeta(\tau) + \sum_{i=1}^{k} \lambda_i B_i(\tau), \\
\zeta(0) &= 0,
\end{align*}
\]

where

\[
A(\tau)(\sigma) := \begin{cases} 
-\delta(\tau, \tau - \sigma) - \bar{u}(\tau)(\sigma) & \text{for } (\tau, \sigma) \in \Omega_1, \\
-\delta(\tau + \sigma, \tau) - \bar{u}(\tau)(\sigma) & \text{for } (\tau, \sigma) \in \Omega_2, \\
-\delta(\sigma + \omega, \omega) - \bar{u}(\tau)(\sigma) & \text{for } (\tau, \sigma) \in \Delta_1, \\
-\delta(T, T - \sigma) - \bar{u}(\tau)(\sigma) & \text{for } (\tau, \sigma) \in \Delta_2.
\end{cases}
\]

and

\[
B_i(\tau) := (-u_i(\tau) + \bar{u}(\tau))\bar{x}(\tau) \text{ for every } i = 1, 2, \ldots, k.
\]

Therefore,

\[
\zeta(\omega) = \int_0^\omega e^{\int_s^\omega A(\mu)d\mu} \left( \sum_{i=1}^{k} \lambda_i [-u_i(s) + \bar{u}(s)]\bar{x}(s) \right) ds.
\]

It is clear that there exists a positive real \(x_{\text{min}} > 0\) such that \(\bar{x}(\tau)(\sigma) \geq x_{\text{min}}\) for all \((\tau, \sigma) \in [0, \omega] \times [-\omega, T]\). Let \(\kappa_{\text{min}}\) be a lower bound for \(e^{\int_s^\omega A(\mu)d\mu}\) for all \(s \in [0, \omega]\).

Now let \(\chi\) be an arbitrary function in \(L_\infty[-\omega, T]\) such that \(\chi(\sigma) \leq 0\) for almost every \(\sigma\) in \([0, T - \omega]\). Then the constant function

\[
w(\tau) \equiv w := \frac{\chi}{\int_0^\omega e^{\int_s^\omega A(\mu)d\mu}.\bar{x}(s)ds}
\]

is measurable and the “variation” \(\zeta(\omega)\) (from (4.6) with \(k = 1\)) corresponding to the admissible control \(u(\tau) = \bar{u}(\tau) - w(\tau)\) coincides with \(\chi(\cdot)\) almost everywhere in \([0, T - \omega]\). Thus the set \(\mathcal{P} - S\) has nonempty interior in \(L_\infty[-\omega, T]\), and, hence, the assumption \((H_3)\) is satisfied. Thus all hypotheses of the main theorem are verified. Therefore, the Pontryagin maximum principle holds true in this example in the presence of the terminal constraint (4.2).

REFERENCES


