Control and Cybernetics

vol. **37** (2008) No. 2

Optimal control of delay-differential inclusions with multivalued initial conditions in infinite dimensions*

by

Boris S. Mordukhovich[1], **Dong Wang**[2] **and Lianwen Wang**[3]

[1] Department of Mathematics, Wayne State University, Detroit, MI, 48202, USA

[2] Department of Mathematics and Computer Science, Fayetteville State University, Fayetteville, NC, 28301, USA

[3] Department of Mathematics and Computer Science, University of Central Missouri, Warrensburg, MO, 64093, USA

e-mail: boris@math.wayne.edu, lwang@ucmo.edu, dwang@uncfsu.edu

Abstract: This paper is devoted to the study of a general class of optimal control problems described by delay-differential inclusions with infinite-dimensional state spaces, endpoints constraints, and multivalued initial conditions. To the best of our knowledge, problems of this type have not been considered in the literature, except for some particular cases when either the state space is finite-dimensional or there is no delay in the dynamics. We develop the method of discrete approximations to derive necessary optimality conditions in the extended Euler-Lagrange form by using advanced tools of variational analysis and generalized differentiation in infinite dimensions. This method consists of the three major parts: (a) constructing a well-posed sequence of discrete-time problems that approximate in an appropriate sense the original continuous-time problem of dynamic optimization; (b) deriving necessary optimality conditions for the approximating discrete-time problems by reducing them to infinite-dimensional problems of mathematical programming and employing then generalized differential calculus; (c) passing finally to the limit in the obtained results for discrete approximations to establish necessary conditions for the given optimal solutions to the original problem. This method is fully realized in the delay-differential systems under consideration.

Keywords: optimal control, variational analysis, endpoint constraints, delay-differential inclusions, multivalued initial conditions, extended Euler-Lagrange formalism, discrete approximations, generalized differentiation, Banach and Asplund spaces, necessary optimality conditions.

*Submitted: January 2008; Accepted: May 2008

1. Introduction, problem formulation, and discussions

The primary objective of this paper is to study a general class of optimal control problems governed by constrained delay-differential inclusions in infinite-dimensional spaces. The main problem of our study is the *generalized Bolza problem* (P) governed by *delay-differential inclusions* in *infinite dimensions* with *endpoint constraints* and *multivalued initial conditions* formulated as follows.

Let X be a Banach *state space*, let $[a,b] \subset I\!R$ be a fixed *time interval*, and let $x\colon [a-\Delta,b] \to X$ be a feasible trajectory/arc of the *delay-differential inclusion*

$$\dot{x}(t) \in F(x(t), x(t-\Delta), t) \quad a.e.\ t \in [a,b], \tag{1.1}$$

$$x(t) \in C(t) \quad a.e.\ t \in [a-\Delta, a), \tag{1.2}$$

$$(x(a), x(b)) \in \Omega \subset X^2 \tag{1.3}$$

with a given time delay $\Delta > 0$, where $F\colon X \times X \times [a,b] \rightrightarrows X$ and $C\colon [a-\Delta, a] \rightrightarrows X$ are set-valued mappings defining the *system dynamics* and the *initial state conditions*, respectively, and where the set $\Omega \subset X^2$ defines the *endpoint constraints*. By a *feasible arc* above we mean a mapping $x\colon [a-\Delta,b] \to X$ that is summable on $[a-\Delta,a]$, Fréchet differentiable for a.e. $t \in [a,b]$, satisfying the *Newton-Leibniz formula*

$$x(t) = x(a) + \int_a^t \dot{x}(s)\,ds \quad \text{for all} \quad t \in [a,b] \tag{1.4}$$

and all the constraints in (1.1)–(1.3), where the integral in (1.4) is taken in the *Bochner sense*. It is well known that for $X = I\!R^n$ the a.e. Fréchet differentiability and Newton-Leibnitz requirements on $x(t), a \le t \le b$, can be equivalently replaced by its *absolute continuity* in the standard sense. In fact, there is a full description of Banach spaces, where this equivalence holds true: they are spaces satisfying the so-called *Radon-Nikodým property* (RNP); see, e.g., Diestel and Uhl (1977). The latter property is fulfilled, in particular, in any reflexive space.

Given now the *endpoint/Mayer* cost function $\varphi\colon X \times X \to I\!R$ and the *integrand/Lagrangian* $f\colon X \times X \times X \times [a,b] \to I\!R$, we consider the *Bolza functional*

$$J[x] := \varphi(x(a), x(b)) + \int_a^b f(x(t), x(t-\Delta), \dot{x}(t), t)\,dt \tag{1.5}$$

and formulate the *dynamic optimization/optimal control* problem (P) as

$$\text{minimize} \quad J[x] \quad \text{subject to} \quad (1.1)-(1.3) \tag{1.6}$$

over feasible arcs $x\colon [a-\Delta,b] \to X$ assuming that $J[x] > -\infty$ for all the feasible arcs and there is at least one feasible $x(\cdot)$ with $J[x] < \infty$.

Note that the generalized Bolza problem (P) unifies a number of particular problems of dynamic optimization (of Mayer type, of Lagrange type, etc.) and contains conventional parameterized forms of optimal control problems governed by *controlled delay-differential equations* of the type

$$\dot{x}(t) = g(x(t), x(t-\Delta), u, t), \quad u \in U, \quad a.e.\ t \in [a, b]. \tag{1.7}$$

Besides other advantages and, of course, higher level of generality of model (1.1) in comparison with that of (1.7), the direct inclusion description (1.1) allows us to cover the *closed-loop* case $U = U(x)$ in (1.7), which is among the most challenging in control theory and the most important for applications. Note also that the presence of the *set-valued* mapping $C(\cdot)$ defined on the initial time interval $[a-\Delta, a)$ in (1.2) is a *specific feature of delay-differential systems* providing an additional source for optimizing the cost functional (1.5) by a choice of the initial condition $x(t) \in C(t)$ on $[a-\Delta, a)$.

The problem (P) under consideration has been studied by Mordukhovich and L. Wang (2003) in the case of finite-dimensional state spaces $X = I\!R^n$; see also the references therein for previous developments on finite-dimensional delay-differential inclusions as well as the books by Mordukhovich (2006b), Smirnov (2002), and Vinter (2000) for more discussions of various approaches and results on nondelayed counterparts of problem (P) and related finite-dimensional control systems.

We are not familiar with *any results* on necessary optimality conditions for optimal control problems governed by delay-differential inclusions and related control systems with *infinite-dimensional* state spaces, even in the case of fixed initial conditions $C(t) = \{c(t)\}$ in (1.2). On the other hand, there are recent developments by Mordukhovich (2006b, 2007) on infinite-dimensional control systems governed by *nondelayed evolution/differential inclusions* of type (1.1) with $\Delta = 0$. We also refer the reader to related (while different) developments in Mordukhovich and D. Wang (2005) concerning *semilinear evolution inclusions* of the type

$$\dot{x}(t) \in Ax(t) + F(x(t), t), \tag{1.8}$$

where $A\colon X \to X$ is an *unbounded* generator of a C_0-semigroup, and where solutions to (1.8) are understood in the *mild* sense. The main approach used in this paper to derive necessary conditions for optimal solutions to the dynamic optimization problem (P) under consideration is the *method of discrete approximations* suggested and implemented by Mordukhovich (2005) in the case of nondelayed differential inclusions in finite-dimensional spaces. This method was extended in Mordukhovich and L. Wang (2003, 2004) to various classes of hereditary functional-differential inclusions in finite dimensions and then in Mordukhovich (2006b, 2007) and in Mordukhovich and D. Wang (2005) to nondelayed differential and evolution inclusions in infinite-dimensional spaces. The version of the discrete approximation method developed in this paper for the problem (P) under consideration consists of the following *three major parts* each of which is certainly of its own interest:

(a) To construct a sequence of *well-posed discrete approximations* of the given *optimal solution* to the original problem (P) in such a way that the approximating discrete-time problems admit optimal solutions, which *strongly* (a.e. *pointwisely* with respect to derivatives) converge to the designated minimizer for the original problem (P). This part of our method is closely related to *sensitivity analysis* of the continuous-type optimization problem (P) for delay-differential inclusions under consideration with respect to discrete approximations, involves not only *qualitative* but also *quantitative* aspects of finite-difference approximations, and essentially relies on the possibility to strongly (and constructively) approximate *any feasible* trajectory to the delay-differential inclusion by feasible trajectories to its finite-difference counterparts.

(b) To derive *necessary optimality conditions* for approximating *discrete-time problems* arising in the well-posed discrete approximation procedure developed in part (a). For any fixed step of approximation, the discrete-time approximating problems can be reduced to non-dynamic problems of constrained *mathematical programming* formulated in *infinite-dimensional spaces*, since the state space in the original problem (P) is infinite-dimensional. A characteristic feature of each of these mathematical programming problems is a specific structure of the involved constraints that are generated by the dynamic constraints of the original problem (P) in the process of discrete approximations. Due to the essential infinite-dimensional nature of the mathematical programs under consideration and in order to avoid additional assumptions in the subsequent procedure of passing to the limit from discrete approximations, we concentrate on deriving *fuzzy necessary optimality conditions* in the obtained problems of mathematical programming and their discrete-time counterparts. This is done on the basis of advanced tools of *variational analysis* and *generalized differentiation* in *infinite-dimensional spaces.*

(c) The final step in the method of discrete approximations is the *passage to the limit* from the obtained necessary optimality conditions in the approximating problems to derive verifiable *exact/pointbased* necessary conditions for the reference optimal solution to the original problem (P). This step, besides employing and unifying the convergence/stability results of part (a) and the fuzzy optimality conditions of part (b), requires the justification of an appropriate *pointwise convergence* of *adjoint trajectories.* This is also done on the basis of advanced tools of infinite-dimensional variational analysis and *robust* generalized differentiation; see below for more details.

The rest of the paper is organized as follows. In Section 2 we formulate and discuss the (fairly general) *standing assumptions* on the *nonconvex* delay-differential inclusion (1.1) and the initial condition (1.2), then construct a sequence of discrete approximations to (1.1) and (1.2) by delay-difference inclusions and establish, in an arbitrary Banach space setting, a principal result on the *strong* $W^{1,1}$*-approximation* of *any* feasible trajectory to the delay-differential system (1.1) and (1.2) by a sequence of feasible trajectories to the delay-difference inclusions constructed above.

Section 3 concerns discrete approximations of the whole problem (P) dealing not only with the underlying delay-differential inclusion, but also with the endpoint constraints (1.2) and the cost functional (1.5). Assuming a certain *relaxation stability* of the original problem and given the reference optimal solution $\bar{x}(\cdot)$ to (P), we construct a *well-posed* sequence of discrete approximations $\{(P_N)\}, N = 1, 2, \ldots$, to (P) in such a way, that each (P_N) admits an optimal solution $\bar{x}_N(\cdot)$, and the sequence $\{\bar{x}_N(t)\}$, naturally extended to the whole interval $[a - \Delta, b]$, $W^{1,1}$-*strongly converges* to $\bar{x}(t)$ as $N \to \infty$. This result requires appropriate *geometric assumptions* on the Banach state space X in question that hold, in particular, when X is *reflexive*.

In Section 4 we briefly overview the basic constructions of *dual-space generalized differentiation* (normals to sets, coderivatives of set-valued mappings, and subdifferentials of extended-real-valued functions) playing a fundamental role in the subsequent variational analysis and deriving necessary optimality conditions in discrete-time and continuous-time optimization problems under consideration, since both classes are *intrinsically nonsmooth*. We also define and discuss the so-called *sequential normal compactness* (SNC) property of sets, which is automatic in finite dimensions while occurs to be a crucial element of variational analysis in infinite-dimensional spaces.

Section 5 is devoted to deriving *necessary optimality conditions* for the discrete approximation problems constructed in Section 3, which are governed by *delay-difference inclusions* with endpoint constraints in infinite-dimensional spaces. As mentioned above, we reduce these problems to *mathematical programs* in Banach spaces with specific types of constraints containing, in particular, an increasing number of *set/geometric constraints* with possibly empty interiors generated by the discrete-time dynamics. The necessary optimality conditions for such mathematical programs and delay-difference inclusions are obtained in this section in *approximate/fuzzy forms*, in contrast to the exact/pointbased forms as in our earlier developments in finite dimensions. The fuzzy results obtained do not require restrictive assumptions on the initial data and are essentially more convenient for the subsequent passage to the limit while deriving the main results of the paper on necessary optimality conditions for the original delay-differential problem (P) in infinite-dimensional spaces. The device developed to establish these fuzzy necessary conditions for delay-difference inclusions is rather involved in comparison, e.g., with the corresponding results in finite dimensions and/or in the nondelayed case. Our approach is based on using advanced tools of generalized differential calculus, coderivative characterizations of metric regularity, etc.

In Sections 6 and 7 we establish the main results of the paper on *necessary optimality conditions* of the *extended Euler-Lagrange type* for the original *generalized Bolza* problem (P) governed by the constrained *delay-differential inclusions* with *infinite-dimensional* state spaces. The final results obtained in these sections are given in the required *exact/pointbased* forms via the *robust* generalized differential constructions reviewed in Section 4. These conditions are derived by passing to the limit from the fuzzy optimality conditions for the approximating delay-difference problems obtained in Section 5 by using the convergence/stability results for discrete approximations established in Sections 2 and 3. Along with these ingredients, the passage to the limit in the approximating necessary optimality conditions requires a delicate variational analysis on the appropriate convergence of *adjoint arcs*; this is mainly done on the basis of generalized differential calculus and dual coderivative characterizations of Lipschitzian stability.

The major difference between the frameworks and results of Section 6 and Section 7 is that the former addresses the original problem (P) with endpoint constraints of the general *geometric type*, while the latter/last section deals with the particular (and more conventional in dynamic optimization) version of the original problem, where endpoint constraints are given explicitly by *finitely many equalities and inequalities* defined by *Lipschitz continuous* (in particular, smooth) functions. The underlying result of Section 6 establishes the extended Euler-Lagrange necessary optimality conditions for the general problem (P) under the *SNC* assumption on the endpoint constraint set Ω, while the result of Section 7 does *not impose* this assumption or the like on the corresponding constraint set described by Lipschitzian equalities and inequalities.

Our notation is basically standard; see Mordukhovich (2006a,b). Unless otherwise stated, all the spaces considered are Banach with the norm $\|\cdot\|$ and the canonical pairing $\langle\cdot,\cdot\rangle$ between the space in question, say X, and its topological dual X^* whose weak* topology is denoted by w^*. We use the symbols $I\!B$ and $I\!B^*$ to signify the closed unit balls of the space in question and its dual, respectively. Given a set-valued mapping $F\colon X \rightrightarrows X^*$, its *sequential Painlevé-Kuratowski upper/outer limit* at $\bar{x}$ is

$$\operatorname*{Lim\,sup}_{x\to\bar{x}} F(x) := \Big\{x^* \in X^* \Big| \exists \text{ sequences } x_k \to \bar{x},\ x_k^* \xrightarrow{w^*} x^* \text{ with } x_k^* \in F(x_k) \text{ as } k \in I\!N := \{1,2,\ldots\}\Big\}. \tag{1.9}$$

2. Discrete approximations of delay-differential inclusions

The main goal of this section is to construct *well-posed discrete approximations* of the original problem (P) that ensure the *strong convergence* of optimal trajectories in the L^1-norm on the "initial tail" interval $[a-\Delta, a]$ and in the $W^{1,1}$-norm on the main interval $[a,b]$. Such a strong convergence plays a crucial role in the subsequent study of delay-differential inclusions via their discrete approximations.

Let $\bar{x}(\cdot)$ be a feasible trajectory to (1.1) with the initial condition (1.2). We impose the following *standing assumptions* on the set-valued mappings F and C used through the whole paper:

(H1) The mapping $C\colon [a-\Delta,a] \rightrightarrows X$ is compact-valued, uniformly bounded
$$C(t) \subset M_C I\!\!B \quad \text{on} \quad [a-\Delta,a] \quad \text{with some} \quad M_C > 0,$$
and Hausdorff continuous for a.e. $t \in [a-\Delta,a]$.

(H2) There are an open set $U \subset M_C I\!\!B$ and two positive numbers L_F and M_F such that $\bar{x}(t) \in U$ for any $t \in [a,b]$, the sets $F(x,y,t)$ are nonempty and compact for all $(x,y,t) \in U \times (M_C I\!\!B) \times [a,b]$, and the following inclusions

$$F(x,y,t) \subset M_F I\!\!B \quad \text{for all } (x,y,t) \in U \times (M_C I\!\!B) \times [a,b], \tag{2.1}$$
$$F(x_1,y_1,t) \subset F(x_2,y_2,t) + L_F(\|x_1-x_2\| + \|y_1-y_2\|)I\!\!B, \tag{2.2}$$

hold whenever (x_1,y_1), $(x_2,y_2) \in U \times (M_C I\!\!B)$ and $t \in [a,b]$.

(H3) $F(x,y,\cdot)$ is Hausdorff continuous for a.e. $t \in [a,b]$ uniformly in $(x,y) \in U \times (M_C I\!\!B)$.

Note that (2.2) signifies the local *Lipschitz continuity* of $F(\cdot,\cdot,t)$ around $(\bar{x}(t), \bar{x}(t-\Delta))$. To clarify the meaning of (H3), consider the so-called *averaged modulus of continuity* $\tau[F;h]$ for $F(x,y,t)$ in $t \in [a,b]$ when $(x,y) \in U \times (M_C I\!\!B)$ defined by

$$\tau[F;h] := \int_a^b \sigma(F;t,h)\,dt, \tag{2.3}$$

where $\sigma(F;t,h) := \sup\{\omega(F;x,y,t,h) \mid (x,y) \in U \times (M_C I\!\!B)\}$ with

$$\omega(F;x,y,t,h)$$
$$:= \sup\big\{\text{haus}(F(x,y,t_1), F(x,y,t_2)) \big|\ t_1,t_2 \in [t-h/2, t+h/2] \cap [a,b]\big\},$$

and where $\text{haus}(\cdot,\cdot)$ stands for the standard *Hausdorff metric* on the space of nonempty and compact subsets of X. It follows from the result by Dontchev and Farkhi (1989) (given in finite dimensions, while their proof works practically without change in the infinite-dimensional setting under consideration) that if $F(x,y,\cdot)$ is Hausdorff continuous for a.e. $t \in [a,b]$ uniformly in $(x,y) \in U \times (M_C I\!\!B)$, then $\tau[F;h] \to 0$ as $h \to 0$. Of course, a simplified version of the above definition applies to the average modulus of continuity $\tau[C;h]$ of the multifunction $C(\cdot)$ on $[a-\Delta,a]$.

Let us now construct a *discrete approximation* of the delay-differential inclusion (1.1) by replacing the time-derivative in (1.1) by the *uniform Euler finite difference*:

$$\dot{x}(t) \approx \frac{x(t+h)-x(t)}{h}, \quad h \to 0.$$

To formalize this procedure, for any natural number $N \in I\!\!N$ take $t_j := a + jh_N$ for $j = -N,\dots,k$ and $t_{k+1} := b$, where $h_N := \Delta/N$ and $k \in I\!\!N$ is defined by

$$a + kh_N \le b < a + (k+1)h_N. \tag{2.4}$$

Note that $t_{-N} = a - \Delta$, $t_0 = a$, and $h_N \to 0$ as $N \to \infty$. Then the sequence of *delay-difference inclusions* approximating (1.1) is constructed as follows:

$$\begin{cases} x_N(t_{j+1}) \in x_N(t_j) + h_N F(x_N(t_j), x_N(t_j - \Delta), t_j) & \text{for } j = 0, \ldots, k, \\ x_N(t_j) \in C(t_j) \quad \text{for } j = -N, \ldots, -1. \end{cases} \tag{2.5}$$

The collection of vectors $\{x_N(t_j) \mid j = -N, \ldots, k+1\}$ satisfying (2.5) is called a *discrete trajectory.* The corresponding collection

$$\Big\{\frac{x_N(t_{j+1}) - x_N(t_j)}{h_N}\Big|\ j = 0, \ldots, k\Big\}$$

is called a *discrete velocity.* We also consider the *extended discrete velocities* defined by

$$v_N(t) := \frac{x_N(t_{j+1}) - x_N(t_j)}{h_N}, \quad t \in [t_j, t_{j+1}),\ j = 0, \ldots, k.$$

It follows from the definition of the Bochner integral that the corresponding *extended discrete trajectories* are given by

$$x_N(t) = x(a) + \int_a^t v_N(s)ds, \qquad t \in [a, b],$$

on the main interval $[a, b]$ and by

$$x_N(t) := x_N(t_j), \quad t \in [t_j, t_{j+1}),\ j = -N, \ldots, -1,$$

on the initial tail interval $[a - \Delta, a)$. Observe that $\dot{x}_N(t) = v_N(t)$ for a.e. $t \in [a, b]$.

The next theorem ensures the *strong approximation* of *any feasible trajectory* $\bar{x}(\cdot)$ to the original delay-differential inclusion given in (1.1) and (1.2) by extended feasible trajectories to its delay-difference counterpart (2.5) in the following sense: the approximation/convergence in the $W^{1,1}([a, b]; X)$-*norm*

$$|x(\cdot)|_{W^{1,1}} := \max_{t \in [a,b]} \|x(t)\| + \int_a^b \|\dot{x}(t)\|\, dt$$

on the main interval $[a, b]$ and the one in the $L^1([a - \Delta, a]; X)$-norm on the initial tail interval $[a - \Delta, a]$. Note that the state space X in Theorem 1 is *arbitrary Banach* and that the strong $W^{1,1}$-convergence of extended discrete trajectories on $[a, b]$ implies the not only their *uniform convergence* on this interval but also the *a.e. pointwise* convergence of their *derivatives* on $[a, b]$ along some subsequence of $\{N\}$ as $N \to \infty$.

THEOREM 1 **(strong approximation by discrete trajectories).** *Let $\bar{x}(\cdot)$ be a feasible trajectory to* (1.1) *and* (1.2) *under assumptions* (H1)–(H3)*, where X is an arbitrary Banach space. Then there is a sequence of solutions $\{z_N(t_j) \mid j = -N, \ldots, k+1\}$ to the delay-difference inclusions* (2.5) *with $z_N(t_0) = \bar{x}(a)$ such that the extended discrete trajectories $z_N(t)$, $t \in [a - \Delta, b]$, converge to $\bar{x}(\cdot)$ strongly in L^1 on $[a - \Delta, a]$ and strongly in $W^{1,1}$ on $[a, b]$ as $N \to \infty$.*

Proof. Given a feasible solution $\bar{x}(t)$, $t \in [a-\Delta, b]$, to the delay-differentiable system (1.1) and (1.2), we have that $\bar{x}(\cdot) \in L^1([a-\Delta, a]; X)$ and that $\bar{x}(\cdot)$ satisfies the Newton-Leibniz formula (1.4). Hence $\bar{x}(\cdot)$ and $\dot{\bar{x}}(\cdot)$ are *strongly measurable* on $[a-\Delta, a]$ and $[a, b]$, respectively. Therefore, rearranging the mesh points t_j, if necessary, we can find a sequence of *simple/step mappings* $\{w_N(\cdot)\}$ on $[a-\Delta, b]$ with $w_N(a) = \bar{x}(a)$ such that each $w_N(\cdot)$ is constant on the intervals $[t_j, t_{j+1})$ as $j = -N, \ldots, k$, that $w_N(\cdot) \to \bar{x}(\cdot)$ on $[a-\Delta, a]$ in the norm of $L^1([a-\Delta, a]; X)$, and that $w_N(\cdot) \to \dot{\bar{x}}(\cdot)$ on $[a, b]$ in the norm of $L^1([a, b]; X)$ as $N \to \infty$. In the estimates below we use the sequence

$$\xi_N := \int_{a-\Delta}^{a} \|\bar{x}(t) - w_N(t)\|\, dt + \int_{a}^{b} \|\dot{\bar{x}}(t) - w_N(t)\|\, dt \to 0 \quad \text{as } N \to \infty. \tag{2.6}$$

Combining the afore-mentioned convergence of $\{w_N(\cdot)\}$ with assumptions (H1)–(H3), we easily find a constant $M > 0$ such that

$$\int_{a-\Delta}^{b} \|w_N(t)\|\, dt \le M \quad \text{for all} \quad N \in I\!N.$$

Further, for each $N \in I\!N$ define the collection $\{u_N(t_j) \mid j = -N, \ldots, k+1\}$ by

$$\begin{cases} u_N(t_j) := w_N(t_j), \quad j = -N, \ldots, 0, \\ u_N(t_{j+1}) := u_N(t_j) + h_N w_N(t_j), \quad j = 0, \ldots, k. \end{cases} \tag{2.7}$$

The corresponding extensions of (2.7) to the intervals $[a-\Delta, a)$ and $[a, b]$ are given by

$$\begin{cases} u_N(t) = w_N(t), \quad t \in [t_j, t_{j+1}),\ j = -N, \ldots, -1, \\ u_N(t) = \bar{x}(a) + \displaystyle\int_a^t w_N(s)\, ds, \quad t \in [a, b]. \end{cases} \tag{2.8}$$

To proceed, we observe that the Lipschitzian condition (2.2) can be equivalently written via the *distance function* on X as

$$\operatorname{dist}(w; F(x_1, y_1, t)) \le \operatorname{dist}(w; F(x_2, y_2, t)) + L_F(\|x_1 - x_2\| + \|y_1 - y_2\|)$$

whenever $w \in X$, $x_1, x_2 \in U$, $y_1, y_2 \in M_C I\!B$, and $t \in [a, b]$. Furthermore, we always have

$$\operatorname{dist}(w; F(x, y, t_1)) \le \operatorname{dist}(w; F(x, y, t_2)) + \operatorname{haus}(F(x, y, t_1), F(x, y, t_2))$$

for any $w \in X$, $x \in U$, $y \in M_C I\!B$ and $t_1, t_2 \in [a, b]$. Using now the *average modulus of continuity* (2.3), we get the relationships

$$\alpha_N := h_N \sum_{j=-N}^{-1} \operatorname{dist}(w_N(t_j); C(t_j))$$
$$+ h_N \sum_{j=0}^{k} \operatorname{dist}(w_N(t_j); F(u_N(t_j), u_N(t_j - \Delta), t_j))$$

$$
\begin{aligned}
&= \sum_{j=-N}^{-1} \int_{t_j}^{t_{j+1}} \operatorname{dist}(w_N(t_j); C(t_j))\, dt \\
&\quad + \sum_{j=0}^{k} \int_{t_j}^{t_{j+1}} \operatorname{dist}(w_N(t_j); F(u_N(t_j), u_N(t_j-\Delta), t_j))\, dt \\
&\le \sum_{j=-N}^{-1} \int_{t_j}^{t_{j+1}} \operatorname{dist}(w_N(t_j); C(t))\, dt \\
&\quad + \sum_{j=0}^{k} \int_{t_j}^{t_{j+1}} \operatorname{dist}(w_N(t_j); F(u_N(t_j), u_N(t_j-\Delta), t))\, dt \\
&\quad + \tau[C; h_N] + \tau[F; h_N], \quad N \in I\!N.
\end{aligned}
$$

Taking the constructions of $\{w_N(\cdot), u_N(\cdot)\}$ and the above estimates into account, we arrive at

$$
\begin{aligned}
\alpha_N &\le (1 + 2L_F) \sum_{j=-N}^{-1} \int_{t_j}^{t_{j+1}} \|w_N(t_j) - \bar{x}(t)\|\, dt \\
&\quad + \sum_{j=0}^{k} \int_{t_j}^{t_{j+1}} \Big(\|w_N(t_j) - \dot{\bar{x}}(t)\| + 2L_F \xi_N \Big)\, dt + \tau[C; h_N] + \tau[F; h_N] \\
&\le (1 + 2L_F)\xi_N + 2L_F \xi_N (b-a) + \tau[C; h_N] + \tau[F; h_N], \quad N \in I\!N,
\end{aligned}
$$

and conclude that $\alpha_N \to 0$ as $N \to \infty$ due to $\xi_N \to 0$ by (2.6), $\tau(C; h_N) \to 0$ by (H1), and $\tau[F; h_N] \to 0$ by (H3).

To continue the proof of the theorem, note that the collections $\{u_N(t_j)\}$ built upon $\{w_N(t_j)\}$ in (2.7) may *not be trajectories* to the delay-difference inclusions (2.5). Let us correct them in such a way that the resulting collections $z_N(t_j)$ satisfy (2.5) and possess the convergence properties stated in the theorem. We construct the desired trajectories $\{z_N(t_j) \mid j = -N, \ldots, k+1\}$ for all $N \in I\!N$ by using the following *proximal algorithm*:

$$
\left\{
\begin{array}{l}
z_N(t_j) = v_{N_j} \ \text{ with } \ v_{N_j} \in C(t_j), \\
\|v_{N_j} - w_N(t_j)\| = \operatorname{dist}(w_N(t_j); C(t_j)), \quad j = -N, \ldots, -1, \\
z_N(t_0) = \bar{x}(a), \\
z_N(t_{j+1}) = z_N(t_j) + h_N v_{N_j} \ \text{ with } \ v_{N_j} \in F(z_N(t_j), z_N(t_j-\Delta), t_j) \\
\text{and } \|v_{N_j} - w_N(t_j)\| = \operatorname{dist}(w_N(t_j); F(z_N(t_j), z_N(t_j-\Delta), t_j)), \\
j = 0, \ldots, k.
\end{array}
\right.
\tag{2.9}
$$

Obviously, $z_N(\cdot)$ in (2.9) are feasible trajectories to (2.5). Now following the proof of Theorem 6.4 in Mordukhovich (2006b) and adapting it to the case of the delay-differential inclusions (1.1) with the set-valued initial conditions (1.2) under consideration, we show that the extensions $z_N(t)$, $t \in [a-\Delta, b]$, of the above discrete trajectories converge to $\bar{x}(t)$ in the L^1-norm topology on $[a-\Delta, a]$ and in the $W^{1,1}$-norm topology on $[a, b]$. The proof is complete. ■

3. Discrete approximations of the generalized Bolza problem

Our next step is to construct a sequence of well-posed discrete approximations of the generalized Bolza problem (P) governed by delay-differential inclusions in such a way that optimal solutions to discrete approximation problems *strongly converge* in the sense specified below to a given optimal solution to the original problem (P).

Let us fix an optimal solution $\bar{x}(t)$, $a-\Delta \le t \le b$, to problem (P), and let $\{z_N(t)\}$, $a-\Delta \le t \le b$, be the sequence of the extended trajectories to the delay-difference inclusions (2.5) approximating $\bar{x}(\cdot)$ in the sense of Theorem 1. Denoting

$$\eta_N := \max_{t\in[a,b]} \|z_N(t)-\bar{x}(t)\| \to 0 \ \text{ as } \ N\to\infty,$$

we construct the sequence of discrete approximation problems (P_N) as follows:

$$\begin{aligned}\text{minimize } J_N[x_N] &:= \varphi(x_N(a),x_N(b)) + \|x_N(a)-\bar{x}(a)\|^2\\ &+\sum_{j=-N}^{-1}\int_{t_j}^{t_{j+1}} \|x_N(t_j)-\bar{x}(t)\|^2 dt\\ &+h_N\sum_{j=0}^{k} f\Big(x_N(t_j),x_N(t_j-\Delta),\frac{x_N(t_{j+1})-x_N(t_j)}{h_N},t_j\Big)\\ &+\sum_{j=0}^{k}\int_{t_j}^{t_{j+1}}\Big\|\frac{x_N(t_{j+1})-x_N(t_j)}{h_N}-\dot{\bar{x}}(t)\Big\|^2 dt\end{aligned} \tag{3.1}$$

subject to the constraints

$$x_N(t_{j+1}) \in x_N(t_j)+h_N F(x_N(t_j),x_N(t_j-\Delta),t_j),\ j=0,\dots,k, \tag{3.2}$$

$$x_N(t_j)\in C(t_j),\ \ j=-N,\dots,-1, \tag{3.3}$$

$$(x_N(a),x_N(b))\in \Omega_N := \Omega+\eta_N I\!B, \tag{3.4}$$

$$\|x_N(t_j)-\bar{x}(t_j)\|\le \epsilon, \quad j=1,\dots,k+1, \tag{3.5}$$

where $\epsilon>0$ is a small given number. In addition to the standing assumptions (H1)–(H3) on (C,F) in (1.1) and (1.2) with some neighborhood U of $\bar{x}(t)$, $a\le t\le b$, we impose the following hypotheses on the behavior of φ, f, and Ω around the optimal trajectory $\bar{x}(\cdot)$ under consideration:

(H4) The cost function φ is continuous on $U\times U$, the constraint set $\Omega\subset X\times X$ is locally closed around $(\bar{x}(a),\bar{x}(b))$, and for some $\nu>0$ the intersection set $\text{proj}_1\Omega\cap(\bar{x}(a)+\nu I\!B)$ is compact in X, where $\text{proj}_1\Omega$ stands for the projection of Ω on the first space X in the product $X\times X$.

(H5) The integrand $f(x,y,v,\cdot)$ is continuous for a.e. $t \in [a,b]$ and bounded uniformly with respect to $(x,y,v) \in U \times (M_C I\!B) \times (M_F I\!B)$; furthermore, there is $\mu > 0$ such that $f(\cdot,\cdot,\cdot,t)$ is continuous on the set

$$A_\mu(t) = \big\{(x,y,v) \in U \times (M_C I\!B) \times (M_F + \mu) I\!B \big|\ v \in F(x,y,s) \text{ for some } s \in (t-\mu, t]\big\}$$

uniformly in $t \in [a,b]$.

Along with the original problem (P), we consider its "relaxed" counterpart constructed in the way well understood in optimal control and variational analysis; see, e.g., the books by Mordukhovich (2006b), Tolstonogov (2000), and Warga (1972). Roughly speaking, the relaxed problem is obtained from (P) by a *convexification* procedure with respect to the *velocity* variable. Let

$$f_F(x,y,v,t) := f(x,y,v,t) + \delta(v; F(x,y,t)),$$

where $\delta(\cdot;\Theta)$ stands for the *indicator function* of the set in question equal to 0 on Θ and to ∞ otherwise. Denote by $\widehat{f}_F(x,y,v,t)$ the *biconjugate* (second conjugate) function to f_F, i.e.,

$$\widehat{f}_F(x,y,v,t) := (f_F)^{**}_v(x,y,v,t).$$

The *relaxed generalized Bolza problem* (R) for the original problem (P) governed by the delay-differential inclusions under consideration is defined as follows:

$$\text{minimize } \widehat{J}[x] := \varphi(x(a), x(b)) + \int_a^b \widehat{f}_F(x(t), x(t-\Delta), \dot{x}(t), t)\, dt \tag{3.6}$$

over feasible trajectories $x(t)$, $a - \Delta \le t \le b$, of the same class as for (P) but to the *convexified* delay-differential inclusion

$$\dot{x}(t) \in \operatorname{clco} F(x(t), x(t-\Delta), t) \quad \text{a.e. } t \in [a,b] \tag{3.7}$$

with the initial condition (1.2) and the endpoint constraints (1.3). As usual, the symbol "clco" in (3.7) stands for the *convex closure* of the set in question.

Denoting by $\inf(P)$ and $\inf(R)$ the *optimal/infimal values* of the cost functionals in problem (P) and (R), respectively, we clearly have $\inf(R) \le \inf(P)$. If

$$\inf(P) = \inf(R),$$

the original problem (P) is said to be *stable with respect to relaxation.* This property, which obviously holds under the *convexity* assumptions with respect to velocity, turns out also to be natural for broad classes of *nonconvex* problems governed by delay-differential inclusions due to the inherent *hidden convexity* of such systems, related, in fact, to the convexity of integrals for set-valued mappings over nonatomic measures; see the afore-mentioned books by Mordukhovich, Tolstonogov, and Warga for more results, discussions, and references.

The next theorem justifies the *existence* of optimal solutions $\bar{x}_N(\cdot)$ to the discrete approximation problems (P_N) and their *strong convergence* to the reference optimal solution $\bar{x}(\cdot)$ to the original problem (P). The strong convergence $\bar{x}_N(\cdot) \to \bar{x}(\cdot)$ is understood in the same sense as in Theorem 1, i.e., as the norm convergence in L^1 on the initial tail interval $[a-\Delta, a]$ and as the norm convergence in $W^{1,1}$ on the main interval $[a,b]$. In fact, under the assumptions made in (H1) and (H2), the strong convergence above can be equivalently replaced by that in the norm of L^p on $[a-\Delta, a]$ and in the norm of $W^{1,p}$ on $[a,b]$ for any $p \geq 1$. We use this in what follows.

In contrast to Theorem 1 held in the general Banach state space X, the main part (ii) of Theorem 2 established below requires additional geometric assumptions imposed on the Banach space X in question. Namely, we assume that *both spaces X and X^* are Asplund*, which automatically holds if X is *reflexive*. Recall that a Banach space X is *Asplund* if every separable subspace of X has a separable dual. This is a broad class of Banach spaces, well investigated in geometric theory and widely applied to many aspects of variational analysis and generalized differentiation; see the books by Borwein and Zhu (2005), Diestel and Uhl (1977), and Mordukhovich (2006a,b) for more details, numerous results, and discussions. Note a remarkable fact from the geometric theory of Banach spaces: X is Asplund *if and only if* the dual space X^* has the Radon-Nikodým property.

Furthermore, part (ii) of the next theorem requires additional assumptions on the initial data in the case of *set-valued* initial conditions (1.2):

(H6) *either* the set $C(t)$ is a singleton $\{c(t)\}$ for a.e. $t \in [a-\Delta, a]$; *or* the set $C(t)$ is convex for a.e. $t \in [a-\Delta, a]$, the mapping $F(x,y,t)$ is linear in y for a.e. $t \in [a, a+\Delta]$, and the function $f(x,y,v,t)$ is convex in (y,v) for a.e. $t \in [a, a+\Delta]$.

THEOREM 2 **(strong convergence of discrete optimal solutions).** *Let $\bar{x}(\cdot)$ be the given optimal solution to the original Bolza problem (P) with the Banach state space X, let $\{(P_N)\}$ as $N \in \mathbb{N}$ be a sequence of discrete approximation problems constructed above, and let the basic assumptions* (H1)–(H5) *be satisfied. Then the following assertions hold:*

(i) *For all $N \in \mathbb{N}$ sufficiently large the problem (P_N) admits an optimal solution.*

(ii) *If, in addition, both spaces X and X^* are Asplund, problem (P) is stable with respect to relaxation, and if, furthermore, the assumptions in* (H6) *are satisfied, then any sequence $\{\bar{x}_N(\cdot)\}$ of optimal solutions to (P_N) extended to the continuous-time interval $[a-\Delta, b]$ converges to $\bar{x}(\cdot)$ as $N \to \infty$ in the L^1-norm topology on $[a-\Delta, a]$ and in the $W^{1,1}$-norm topology on $[a,b]$.*

Proof. To justify (i), we take $\epsilon > 0$ in (3.5) such that $\bar{x}(t) + \epsilon \mathbb{B} \subset U$ for all $t \in [a,b]$ and consider numbers $N \in \mathbb{N}$ so large that $\eta_N < \epsilon$ along the numerical sequence $\{\eta_N\}$ used in the construction of problem (P_N). Note that for such large $N \in I\!N$ each problem (P_N) has *feasible solutions*, since the trajectory $z_N(\cdot)$ from Theorem 1 satisfies all the constraints (3.2)–(3.5). The existence of *optimal solutions* to (P_N) follows now from the classical *Weierstrass existence theorem* due to the compactness and continuity assumptions made in (H1)–(H5).

To justify the *strong convergence* assertion (ii) of the theorem, consider again the sequence $z_N(\cdot)$ that strongly approximates $\bar{x}(\cdot)$ by Theorem 1. Since each $z_N(\cdot)$ is feasible to (P_N), we have

$$J_N[\bar{x}_N] \le J_N[z_N] \quad \text{for all } N \in I\!N.$$

It can be shown similarly to the proof of Theorem 6.13 in Mordukhovich (2006b) given for the case of nondelayed differential inclusions that

$$J_N[z_N] \to J[\bar{x}] \quad \text{as } N \to \infty$$

by (H5) and by using the Lebesgue dominated convergence theorem valid for the Bochner integral in arbitrary Banach spaces. The above two relationships easily yield that

$$\limsup_{N\to\infty} J_N[\bar{x}_N] \le J[\bar{x}] \tag{3.8}$$

under the the general assumptions of the theorem as in assertion (i).

Let us show next that (3.8) implies the strong convergence $\bar{x}_N(\cdot) \to \bar{x}(\cdot)$ claimed in (ii) under the additional assumptions made therein. Due the aforementioned equivalence between the $L^1/W^{1,1}$ and $L^2/W^{1,2}$ convergence in the theorem, we need to prove that

$$\begin{aligned} \rho_N := &\int_{a-\Delta}^{a} \|\bar{x}_N(t) - \bar{x}(t)\|^2\,dt + \|\bar{x}_N(a) - \bar{x}(a)\|^2 \\ &+ \int_a^b \|\dot{\bar{x}}_N(t) - \dot{\bar{x}}(t)\|^2\,dt \to 0 \end{aligned} \tag{3.9}$$

as $N \to \infty$. Arguing by contradiction, suppose that (3.9) does not hold. Then there is a limiting point ρ of $\{\rho_N\}$ such that $\rho > 0$. Suppose, without loss of generality, that $\rho_N \to \rho > 0$ as $N \to \infty$.

To proceed further, observe that under the assumptions made in (ii) *both spaces* X and X^* enjoy the *Radon-Nikodým property* (RNP). Indeed, the one for X^* is equivalent, as mentioned above, to the Asplund property of X, while the Asplund property of X^* ensures the RNP of X due to the latter fact and the inclusion $X \subset X^{**}$. Note also that both sequences $\{\bar{x}_N(t)\}$, $t \in [a - \Delta, a]$, and $\{\dot{\bar{x}}_N(t)\}$, $t \in [a,b]$, are *uniformly bounded* by the assumptions (H1) and (H2). Applying now to these sequences the *Dunford theorem* on the *weak compactness* in, respectively, the spaces $L^1([a-\Delta,b];X)$ and $L^1([a,b];X)$ (see, e.g., Theorem IV.1 in Diestel and Uhl, 1977), we find $\widetilde{x}(\cdot) \in L^1([a-\Delta,b];X)$ and $v(\cdot) \in L^1([a,b];X)$ such that

$$\bar{x}_N(\cdot) \to \widetilde{x}(\cdot) \text{ weakly in } L^1([a-\Delta,a];X), \tag{3.10}$$

$$\dot{\bar{x}}_N(\cdot) \to v(\cdot) \text{ weakly in } L^1([a,b];X) \tag{3.11}$$

along a subsequence of $N \to \infty$, where all $N \in I\!N$ can be taken without loss of generality. Note that each $x_N(\cdot)$, $a \le t \le b$, satisfies the Newton-Leibniz formula

$$\bar{x}_N(t) = \bar{x}_N(a) + \int_a^t \dot{\bar{x}}_N(s)\,ds \ \text{ for all } \ t \in [a,b] \ \text{ and } \ N \in I\!N. \tag{3.12}$$

Furthermore, by compactness of the set $\{\bar{x}_N(a) |\ N \in I\!N\}$ in X due to assumption (H4) and by taking into account the relationships

$$\begin{aligned} \dot{\bar{x}}_N(t) =& \dot{\bar{x}}_N(t_j) \in F(\bar{x}_N(t_j), \bar{x}_N(t_j-\Delta), t_j), \\ & t \in [t_j, t_{j+1}),\ j = 0,\ldots,k,\ N \in I\!N, \end{aligned} \tag{3.13}$$

and assumptions (H2) and (H3) imposed on F, we conclude that the sequence $\{x_N(\cdot)\}$ contains a *convergent subsequence* in the *norm topology* of $C([a,b];X)$; see, e.g., Theorem 3.4.2 in Tolstonogov (2000). Since the Bochner integral is well known to be *weakly continuous* as an operator from $L^1([a,b];X)$ to X, we pass to the limit in (3.12) as $N \to \infty$ and deduce from (3.11) and from the above discussions that there is $\widetilde{x}(\cdot) \in C([a,b];X)$ such that

$$\widetilde{x}(t) = \widetilde{x}(a) + \int_a^t v(s)\,ds \ \text{ for all } \ t \in [a,b], \tag{3.14}$$

which immediately implies the absolute continuity and a.e. Fréchet differentiability of $\widetilde{x}(\cdot)$ on $[a,b]$ with $v(t) = \dot{\widetilde{x}}(t)$ for a.e. $t \in [a,b]$.

Thus now we have the arc $\widetilde{x}\colon [a-\Delta,b] \to X$ built in (3.10) and (3.14) with $v(\cdot)$ constructed in (3.11). Observe, first of all, that $\widetilde{x}(\cdot)$ is a *feasible arc* to the *relaxed* problem (R). Indeed, the classical *Mazur theorem* in functional analysis allows us to conclude from the weak convergence in (3.10) and (3.11) that there are *convex combinations* of elements from $\{\bar{x}_N(t)\}$, $t \in [a-\Delta,a]$, and $\{\dot{\bar{x}}_N(t)\}$, $t \in [a,b]$, which converge to $\widetilde{x}(t)$ and $v(t) = \dot{\widetilde{x}}(t)$ *strongly* in $L^1([a-\Delta,a];X)$ and $L^1([a,b];X)$, respectively. Hence, some subsequences of these convex combinations (as usual we take the whole sequences without loss of generality) converge *almost pointwisely* to $\widetilde{x}(t)$ and $\dot{\widetilde{x}}(t)$ on the corresponding intervals. This immediately implies by passing to the limit in (3.3) for $\bar{x}_N(\cdot)$ due to (H1) and the assumed *convexity* of the sets $C(t)$ in (H6) that $\widetilde{x}(t)$ satisfies (1.2), which is the initial condition for the relaxed problem (R). The fulfillment of the endpoint constraints (1.3) for $\widetilde{x}(\cdot)$ can be easily justified by passing to the limit in (3.4) and taking into account that $\eta_N \to 0$ as $N \to \infty$. Passing finally to the limit in the discrete inclusions (3.13) and employing the a.e. pointwise convergence of the convex combinations of $\{\bar{x}_N(\cdot)\}$ on $[a-\Delta,a]$ and $\{\dot{\bar{x}}_N(\cdot)\}$ on $[a,b]$ justified above as well as the *convexified* structure of the relaxed delay-differential inclusion (3.7) under the *linearity* assumption on F in (H6) in the case of the *multivalued* initial tail mapping $C(\cdot)$, we conclude that $\widetilde{x}(\cdot)$ satisfies (3.7) on $[a,b]$, and hence it is a feasible arc to (R).

Let us further proceed with the passage to the limit in the cost functional (3.1) along the sequence $\{\bar{x}_N(\cdot)\}$. By the identity

$$\begin{aligned} &h_N\sum_{j=0}^{k} f\Big(\bar{x}_N(t_j),\bar{x}_N(t_j-\Delta),\frac{\bar{x}_N(t_{j+1})-\bar{x}_N(t_j)}{h_N},t_j\Big)\\ &=\sum_{j=0}^{k}\int_{t_j}^{t_{j+1}} f(\bar{x}_N(t_j),\bar{x}_N(t_j-\Delta),\dot{\bar{x}}_N(t_j),t_j)\,dt, \end{aligned}$$

by structure (3.6) of the integrand in (R), by assumptions (H5) and (H6) on f, and by the a.e. pointwise convergence of the convex combinations above we get

$$\begin{aligned} &\int_a^b \widehat{f}_F(\widetilde{x}(t),\widetilde{x}(t-\Delta),\dot{\widetilde{x}}(t),t)\,dt\\ &\le \liminf_{N\to\infty} h_N\sum_{j=0}^{k} f\Big(\bar{x}_N(t_j),\bar{x}_N(t_j-\Delta),\frac{\bar{x}_N(t_{j+1})-\bar{x}_N(t_j)}{h_N},t_j\Big). \end{aligned}$$

Observe further that the integral functionals

$$I_1[v]:=\int_{a-\Delta}^{a}\|v(t)-\bar{x}(t)\|^2dt \quad\text{and}\quad I_2[v]:=\int_a^b\|v(t)-\dot{\bar{x}}(t)\|^2dt$$

are *lower semicontinuous* in the *weak topology* of $L^1([a-\Delta,a];X)$ and $L^1([a,b];X)$, respectively, due to the *convexity* of the integrands therein in v. Since

$$\begin{aligned} &\sum_{j=-N}^{-1}\int_{t_j}^{t_{j+1}}\|\bar{x}_N(t_j)-\bar{x}(t)\|^2dt=\int_{a-\Delta}^{a}\|\bar{x}_N(t)-\bar{x}(t)\|^2dt \quad\text{and}\\ &\sum_{j=0}^{k}\int_{t_j}^{t_{j+1}}\Big\|\frac{\bar{x}_N(t_{j+1})-\bar{x}_N(t_j)}{h_N}-\dot{\bar{x}}(t)\Big\|^2dt=\int_a^b\|\dot{\bar{x}}_N(t)-\dot{\bar{x}}(t)\|^2dt, \end{aligned}$$

the afore-mentioned weak lower semicontinuity implies that

$$\begin{aligned} &\int_{a-\Delta}^{a}\|\widetilde{x}(t)-\bar{x}(t)\|^2dt\le\liminf_{N\to\infty}\sum_{j=-N}^{-1}\int_{t_j}^{t_{j+1}}\|\bar{x}(t_j)-\bar{x}(t)\|^2dt \quad\text{and}\\ &\int_a^b\|\dot{\widetilde{x}}(t)-\dot{\bar{x}}(t)\|^2dt\le\liminf_{N\to\infty}\sum_{j=0}^{k}\int_{t_j}^{t_{j+1}}\Big\|\frac{\bar{x}_N(t_{j+1})-\bar{x}_N(t_j)}{h_N}-\dot{\bar{x}}(t)\Big\|^2dt. \end{aligned}$$

Thus, passing to the limit in (3.1) and taking into account the construction of ρ_N in (3.9) and the *upper estimate* (3.8) established above, we arrive at the inequalities

$$\widehat{J}[\widetilde{x}] + \rho \leq \liminf_{N\to\infty} J_N[\bar{x}_N] \leq \limsup_{N\to\infty} J_N[\bar{x}_N] \leq J[\bar{x}],$$

which imply by the *relaxation stability* of (P) and by assumed *positivity* of $\rho > 0$ that

$$\widehat{J}[\widetilde{x}] < J[\bar{x}] = \widehat{J}[\bar{x}]. \tag{3.15}$$

Since the arc $\widetilde{x}(\cdot)$ is proved to be *feasible* to the relaxed problem (R), the strict inequality in (3.15) obviously *contradicts* the *optimality* of $\bar{x}(\cdot)$ to (R) and to (P), and hence the positivity assumption on ρ was *wrong*. Thus we get the convergence $\rho_N \to 0$ in (3.9), which justifies the *strong convergence* of optimal solutions $\bar{x}_N(\cdot) \to \bar{x}(\cdot)$ as $N \to \infty$ claimed in assertion (ii). This completes the proof of the theorem. ■

4. Tools of generalized differentiation

Theorem 2 on the strong convergence of discrete approximations *makes a bridge* between the given optimal solution $\bar{x}(\cdot)$ to the original Bolza problem (P) governed by the delay-differential inclusion (1.1) and optimal solutions to its discrete-time counterparts (P_N). This determines our further strategy: to derive first the necessary conditions for the optimal solutions $\bar{x}_N(\cdot)$ to the approximating problems (P_N) and then get those for the given optimal solution $\bar{x}(\cdot)$ to the original problem (P) by passing to the limit from the ones for the discrete approximations.

A characteristic feature of problems (P) and (P_N) is their *intrinsic nonsmoothness* that is inevitably generated by the *dynamic constraints* in (1.1) and (2.5), even in the case of smooth cost functions and endpoint constraints, although we do not restrict our consideration to the smooth data in any of these parts. To deal with nonsmoothness, we use appropriate tools of *generalized differentiation* briefly reviewed in this section based on the book by Mordukhovich (2006a), where the reader can find more details and discussions. We also refer the reader to the recent books by Borwein and Zhu (2005) and by Schirotzek (2007) for related and additional material on generalized differentiation. Since the corresponding constructions and properties are used in this paper in the *Asplund space* framework, their definitions are adjusted to this setting.

We start with generalized normals to nonempty *sets* that are locally closed around the references points. Given $\Omega \subset X$, define the (basic, limiting, Mordukhovich) *normal cone* to Ω at $\bar{x} \in \Omega$ by

$$N(\bar{x};\Omega) := \operatorname*{Lim\,sup}_{x \xrightarrow{\Omega} \bar{x}} \widehat{N}(x;\Omega) \tag{4.1}$$

via the sequential Painlevé-Kuratowski outer/upper limit (1.9) of the *prenormal/Fréchet normal cone* to Ω at $x \in \Omega$ given by

$$\widehat{N}(x;\Omega) := \Big\{x^* \in X^* \Big|\ \limsup_{u \xrightarrow{\Omega} x} \frac{\langle x^*, u - x\rangle}{\|u - x\|} \le 0\Big\}, \tag{4.2}$$

where the symbol $x \xrightarrow{\Omega} \bar{x}$ indicates that $x \to \bar{x}$ with $x \in \Omega$. Note that for convex sets Ω we have

$$N(\bar{x};\Omega) = \widehat{N}(\bar{x};\Omega) = \big\{x^* \in X^* \big|\ \langle x^*, x - \bar{x}\rangle \le 0 \quad \text{for all } x \in \Omega\big\}. \tag{4.3}$$

Given a *set-valued mapping* $F\colon X \rightrightarrows Y$ and a point $(\bar{x},\bar{y}) \in \operatorname{gph} F$, define the *basic coderivative* of F at $(\bar{x},\bar{y})$ and the *Fréchet coderivative* of F at this point by, respectively,

$$D^*F(\bar{x},\bar{y})(y^*) := \big\{x^* \in X^* \big|\ (x^*, -y^*) \in N((\bar{x},\bar{y}); \operatorname{gph} F)\big\}, \tag{4.4}$$

$$\widehat{D}^*F(\bar{x},\bar{y})(y^*) := \big\{x^* \in X^* \big|\ (x^*, -y^*) \in \widehat{N}((\bar{x},\bar{y}); \operatorname{gph} F)\big\}. \tag{4.5}$$

Note that both coderivatives (4.4) and (4.5) are positively homogeneous set-valued mappings from Y^* to X^*. They both are single-valued and linear

$$D^*F(\bar{x})(y^*) = \widehat{D}^*F(\bar{x})(y^*) = \big\{\nabla F(\bar{x})^* y^*\big\} \ \text{ for all } \ y^* \in Y^*$$

if $F\colon X \to Y$ is single-valued and C^1 around $\bar{x}$, or merely strictly differentiable at this point.

Given now an *extended-real-valued function* $\varphi\colon X \to \overline{I\!R} := (-\infty, \infty]$ finite at $\bar{x}$, the (basic, limiting, Mordukhovich) *subdifferential* of φ at $\bar{x}$ is defined by

$$\partial\varphi(\bar{x}) := \operatorname*{Lim\,sup}_{x \xrightarrow{\varphi} \bar{x}} \widehat{\partial}\varphi(x), \tag{4.6}$$

where $x \xrightarrow{\varphi} \bar{x}$ means that $x \to \bar{x}$ with $\varphi(x) \to \varphi(\bar{x})$, and where $\widehat{\partial}\varphi(x)$ stands for the *Fréchet subdifferential* of φ at x defined by

$$\widehat{\partial}\varphi(x) := \Big\{x^* \in X^* \Big|\ \liminf_{u \to x} \frac{\varphi(u) - \varphi(x) - \langle x^*, u - x\rangle}{\|u - x\|} \ge 0\Big\}. \tag{4.7}$$

In this paper, besides the generalized differential constructions defined above, we also use some of their extended limiting versions for *moving* (parameter-dependent) objects needed in the case of nonautonomous systems. Given a moving set $\Omega\colon T \rightrightarrows X$, the *extended normal cone* to $\Omega(\bar{t})$ at $\bar{x} \in \Omega(\bar{t})$ is defined by

$$N_+(\bar{x};\Omega(\bar{t})) := \operatorname*{Lim\,sup}_{(x,t) \xrightarrow{\operatorname{gph}\Omega} (\bar{x},\bar{t})} \widehat{N}(x;\Omega(t)). \tag{4.8}$$

Given a parameter-dependent function $\varphi\colon X\times T\to\overline{I\!R}$ finite at $(\bar x,\bar t)$, the *extended subdifferential* of $\varphi(\cdot,\bar t)$ at $\bar x$ is defined by

$$\partial_+\varphi(\bar x,\bar t)=\mathop{\rm Lim\,sup}_{(x,t)\xrightarrow{\varphi}(\bar x,\bar t)}\widehat\partial\varphi(x,t),\tag{4.9}$$

where $\widehat\partial\varphi(\cdot,t)$ is taken with respect to x under fixed t. Obviously, the extended normal cone (4.8) and the extended subdifferential (4.9) reduce to the basic objects (4.1) and (4.6) if, respectively, $\Omega(\cdot)$ and $\varphi(\cdot,t)$ are independent of t.

It is important to emphasize that the limiting constructions (4.1), (4.4), (4.6), (4.8), and (4.9), being generally *nonconvex*, enjoy *full calculi* in the framework of Asplund spaces, while the Fréchet-like constructions (4.2), (4.5), and (4.7) satisfy "fuzzy calculus rules" used in the next section. All these results are based on the fundamental *extremal/variational principles* of variational analysis.

An important ingredient of variational analysis in infinite-dimensional spaces is the *sequential normal compactness* (SNC) property of sets defined as follows: $\Omega\subset X$ is *SNC* at $\bar x\in\Omega$ if, for any sequences $\{(x_n,x_n^*)\}\subset X\times X^*$ satisfying

$$x_n\xrightarrow{\Omega}\bar x\ \text{ as }\ n\to\infty\ \text{ and }\ x_n^*\in\widehat N(x_n;\Omega)\ \text{ for all }\ n\in I\!N,$$

we have the implication

$$x_n^*\xrightarrow{w^*}0\Longrightarrow\|x_n^*\|\to0\ \text{ as }\ n\to\infty.$$

This property automatically holds if either X is *finite-dimensional*, or Ω is a *convex* set having *nonempty relative interior* with respect to its closed affine hull of *finite codimension*. More generally, Ω enjoys the SNC property at $\bar x$ if it is *compactly epi-Lipschitzian* (CEL) around this point in the sense of Borwein and Strójwas, which is implied in turn by the *epi-Lipschitzian* property in the sense of Rockafellar; see Subsection 1.1.4 of the afore-mentioned book by Mordukhovich (2006a) for more details, references, and discussions. A crucial feature of SNC is *full calculus* (i.e., comprehensive rules ensuring the preservation of this property under various operations), which is also based on the *extremal/variational principles*.

5. Necessary conditions for delay-difference inclusions

In this section we obtain necessary conditions for optimal solutions to the discrete optimization problems (P_N). We reduce these discrete-time dynamic optimization problems to problems of mathematical programming with functional, operator, and *many* geometric constraints. To conduct a local variational analysis of the mathematical programs and discrete optimization problems under consideration, we to use the tools of generalized differentiation, discussed in Section 4.

It is easy to observe that each discrete optimization problem (P_N), for any fixed $N \in I\!N$ and the corresponding number $k \in I\!N$ defined in (2.4), can be equivalently written as the following *mathematical program* (MP):

$$\begin{cases} \text{minimize } \phi_0(z) \quad \text{subject to} \\ \phi_j(z) \le 0,\ j = 1, \ldots, s, \\ g(z) = 0, \\ z \in \Theta_j \subset Z,\ j = 1, \ldots, l, \end{cases} \tag{5.1}$$

in the *infinite-dimensional* space $Z := X^{N+2k+3}$ with respect to the “long” variable

$$\begin{aligned} z &= (x_{-N}^N, \ldots, x_{k+1}^N, y_0^N, \ldots, y_k^N) \\ &:= (x^N(t_{-N}), \ldots, x^N(t_{k+1}), y^N(t_0), \ldots, y^N(t_k)) \in Z \end{aligned} \tag{5.2}$$

subject to, respectively, *inequality* constraints defined by the functions ϕ_j, *operator* constraints defined by the mappings g whose image space is *infinite-dimensional*, and the increasing number of *geometric* constraints defined by the sets Θ_j particularly generated by the *delay-difference dynamics*. The initial data (ϕ_j, g, Θ_j) in (5.1) are given by:

$$\begin{aligned} \phi_0(z) :&= \varphi(x_0^N, x_{k+1}^N) + \|x_0^N - \bar{x}(a)\|^2 + \sum_{j=-N}^{-1} \int_{t_j}^{t_{j+1}} \|x_j^N - \bar{x}(t)\|^2 dt \\ &+ h_N \sum_{j=0}^{k} f(x_j^N, x_{j-N}^N, y_j^N, t_j) + \sum_{j=0}^{k} \int_{t_j}^{t_{j+1}} \|y_j^N - \dot{\bar{x}}(t)\|^2 dt, \end{aligned} \tag{5.3}$$

$$\phi_j(z) := \|x_j^N - \bar{x}(t_j)\| - \epsilon, \quad j = 1, \ldots, k+1, \tag{5.4}$$

$$\begin{aligned} &g(z) = (g_0(z), \ldots, g_k(z)) \ \text{ with} \\ &g_j(z) := x_{j+1}^N - x_j^N - h_N y_j^N, \quad j = 0, \ldots, k, \end{aligned} \tag{5.5}$$

$$\Theta_j := \big\{(x_{-N}^N, \ldots, y_k^N) \mid \ x_j^N \in C(t_j)\big\}, \quad j = -N, \ldots, -1, \tag{5.6}$$

$$\Theta_j := \big\{(x_{-N}^N, \ldots, y_k^N) \mid \ y_j^N \in F(x_j^N, x_{j-N}^N, t_j)\big\}, \quad j = 0, \ldots, k, \tag{5.7}$$

$$\Theta_{k+1} := \big\{(x_{-N}^N, \ldots, y_k^N) \mid (x_0^N, x_{k+1}^N) \in \Omega_N\big\}. \tag{5.8}$$

The next theorem establishes the necessary conditions for optimal solutions to each problem (P_N). In contrast to the case of delay-difference systems with finite-dimensional state spaces as in Mordukhovich and L. Wang (2003), we now obtain optimality conditions in *fuzzy/approximate* discrete-time forms of the *Euler-Lagrange* and *transversality* inclusions, expressed in terms of the Fréchet-like generalized differential constructions reviewed in Section 4. The major reason for this is that the optimality conditions of the fuzzy type for discrete-time systems can be obtained under fairly general and nonrestrictive assumptions on the initial data, which happens to be much more convenient to derive the main results of the paper on necessary optimality conditions for delay-differential inclusions in *infinite dimensions* by passing to the limit from discrete approximations; see Sections 6 and 7. The proof of the fuzzy optimality conditions in the next theorem is largely based on applying the *fuzzy calculus rules* for Fréchet normals, subgradients, and coderivatives and on *dual neighborhood characterizations* of the *Lipschitzian* and *metric regularity* properties for nonsmooth mappings taken from Mordukhovich (2006a). Note that fuzzy calculus rules provide representations of the underlying normals/subgradients/coderivatives of compositions at the reference points via those at points that are arbitrary close to the reference ones. Just for *notational simplicity* and convenience, we suppose in the formulation and proof of the next theorem that these arbitrary close points *reduce* to the reference ones in question. It makes *no difference* for the limiting procedure to derive the main necessary optimality conditions for delay-differential inclusions in what follows.

THEOREM 3 **(approximate Euler-Lagrange conditions for delay-difference inclusions).** *Let $\bar z^N(\cdot)$ be an optimal solution to problem (P_N) with any fixed $N \in I\!N$ sufficiently large under the standing hypotheses* (H1)–(H3). *Denote $F_j := F(\cdot,\cdot,t_j)$ and $f_j := f(\cdot,\cdot,\cdot,t_j)$ and assume, in addition, that X is Asplund and that the functions φ and f_j are Lipschitz continuous around $(\bar x_0^N, \bar x_{k+1}^N)$ and $(\bar x_j^N, \bar x_{j-N}^N, \bar y_j^N)$, respectively, for $j = 0,\ldots,k$. Consider the quantities*

$$\begin{cases} \theta_j^N := 2\displaystyle\int_{t_j}^{t_{j+1}} \Big\|\frac{\bar x_{j+1}^N - \bar x_j^N}{h_N} - \dot{\bar x}(t)\Big\|\,dt, \quad j = 0,\ldots,k,\\ \sigma_j^N := 2\displaystyle\int_{t_j}^{t_{j+1}} \|\bar x_j^N - \bar x(t)\|\,dt, \quad j = -N,\ldots,-1. \end{cases} \tag{5.9}$$

Then there exists a number $\gamma > 0$ independent of N such that for any sequences of positive numbers $\varepsilon_N \to 0$ as $N \to \infty$ there are multipliers $\lambda^N \ge 0$ and sequences of the discrete adjoint arcs $p_j^N \in X^*$ $(j = 0,\ldots,k+1)$, and $q_j^N \in X^*$ $(j = -N,\ldots,k+1)$, satisfying the following relationships:
—the nontriviality condition

$$\lambda^N + \|p_{k+1}^N\| \ge \gamma, \tag{5.10}$$

— the approximate Euler-Lagrange inclusion

$$\begin{cases} \Big(\dfrac{p_{j+1}^N - p_j^N}{h_N}, \dfrac{q_{j-N+1}^N - q_{j-N}^N}{h_N}, -\dfrac{\lambda^N\theta_j^N}{h_N}a_j^N + p_{j+1}^N + q_{j+1}^N\Big)\\ \in \lambda^N\widehat\partial f(\bar x_j^N, \bar x_{j-N}^N, \bar y_j^N, t_j) + \widehat N((\bar x_j^N, \bar x_{j-N}^N, \bar y_j^N); \mathrm{gph}\, F_j)\\ \quad +\varepsilon_N I\!B^* \text{ with some } a_j^N \in I\!B^*,\ j = 0,\ldots,k, \end{cases} \tag{5.11}$$

— *the approximate tail conditions*

$$\begin{cases} -\dfrac{q_{j+1}^N - q_j^N}{h_N} - \lambda^N \dfrac{\sigma_j^N}{h_N} b_j^N \in \widehat{N}(\bar{x}_j^N; C(t_j)) + \varepsilon_N I\!B^* \\ \quad \text{with some } b_j^N \in I\!B^*,\ j = -N, \ldots, -1, \\ q_j^N = 0, \quad j = k - N + 1, \ldots, k + 1, \end{cases} \tag{5.12}$$

— *and the approximate transversality inclusion*

$$(p_0^N + q_0^N, -p_{k+1}^N) \in \lambda^N \widehat{\partial}\varphi(\bar{x}_0^N, \bar{x}_{k+1}^N) + \widehat{N}((\bar{x}_0^N, \bar{x}_{k+1}^N); \Omega_N) + \varepsilon_N I\!B^* \tag{5.13}$$

Proof. Consider problem (P_N) in the equivalent *mathematical programming* form (5.1) for the decision variable $z \in Z$ in (5.2) with the initial data defined in (5.3)–(5.8). Given $\epsilon > 0$ in (3.5), take $N \in I\!N$ so large that constraints (3.5) hold as *strict* inequalities, which is ensured by Theorem 2. Then *all* the inequality constraints in (5.4) are *inactive* at the optimal solution

$$\begin{aligned} \bar{z}^N &:= (\bar{x}_{-N}^N, \ldots, \bar{x}_{k+1}^N, \bar{y}_0^N, \ldots, \bar{y}_k^N) \\ &= (\bar{x}^N(t_{-N}), \ldots, \bar{x}^N(t_{k+1}), \bar{y}^N(t_0), \ldots, \bar{y}^N(t_k)) \end{aligned}$$

to (P_N), and thus the functions ϕ_j, $j = 1, \ldots, k+1$, can be *ignored* in the arguments below.

Let us examine the following two *mutually exclusive cases* in the proof of the theorem, which complement each other.

Case 1. Assume that the operator constraint mapping $g : X^{N+2k+3} \to X^{k+1}$ in (5.5) is *metrically regular* at $\bar{z}^N$ relative to the set

$$\Theta := \bigcap_{j=-N}^{k+1} \Theta_j, \tag{5.14}$$

with Θ_j taken from (5.6)–(5.8), in the sense that there is a constant $\mu > 0$ and a neighborhood V of $\bar{z}^N$ such that the *distance estimate*

$$\operatorname{dist}(z; S) \le \mu \, \|g(z) - g(\bar{z}^N)\| \quad \text{for all} \quad z \in \Theta \cap V$$

with $S := \{ z \in \Theta \mid \ g(z) = g(\bar{z}^N) \}$ is satisfied. Then, by Ioffe's *exact penalization* theorem (see, e.g., Theorem 5.16 in Mordukhovich, 2006b), we conclude that $\bar{z}^N$ is a local optimal solution to the *unconstrained penalized problem*:

$$\text{minimize } \phi_0(z) + \mu \left(\|g(z)\| + \operatorname{dist}(z; \Theta) \right)$$

for all $\mu > 0$ sufficiently large. It easily follows from construction (4.7) of the Fréchet subdifferential that the *Fermat generalized stationary condition*

$$0 \in \widehat{\partial}\big(\phi_0(\cdot) + \mu\|g(\cdot)\| + \mu \operatorname{dist}(\cdot, \Theta)\big)(\bar{z}^N) \tag{5.15}$$

holds. Taking any sequence $\varepsilon_N \downarrow 0$ as $N \to \infty$ and employing in (5.15) the *fuzzy sum rule* and then the formula for computing Fréchet subgradients of the *distance function* from, respectively, Theorem 2.33(b) and Proposition 1.95 in Mordukhovich (2006a), we get

$$0 \in \widehat{\partial}\phi_0(\bar{z}^N) + \sum_{j=0}^{k} \nabla g_j(\bar{z}^N)^* e_j^* + \widehat{N}(\bar{z}^N;\Theta) + \varepsilon_N h_N I\!B^*, \tag{5.16}$$

for some $e_j^* \in X^*$ satisfying

$$\begin{aligned}&\sum_{j=0}^{k} \nabla g_j(\bar{z}^N)^* e_j^*\\ &= \big(0,\ldots,0,-e_0^*,e_0^*-e_1^*,\ldots,e_{k-1}^*-e_k^*,e_k^*,-h_N e_0^*,\ldots,-h_N e_k^*\big)\end{aligned} \tag{5.17}$$

due to the specific structure of the operator constraints (5.5) and the simple chain rule for the composition $\|g(z)\| = (\psi \circ g)(z)$ with $\psi(v) := \|v\|$ and the smooth mapping g from (5.5).

To proceed further, we apply to the set Θ in (5.14) the *fuzzy intersection rule* from Lemma 3.1 in Mordukhovich (2006a) ensuring that

$$\widehat{N}(\bar{z}^N;\Theta) \subset \widehat{N}(\bar{z}^N;\Theta_{-N}) + \ldots + \widehat{N}(\bar{z}^N;\Theta_{k+1}) + \varepsilon_N h_N I\!B^*.$$

Taking into account the sum structure of cost functional ϕ_0 in (5.3) and the specific forms of the terms therein, we get from the afore-mentioned fuzzy sum rule that

$$\begin{aligned}&\widehat{\partial}\phi_0(\bar{z}^N) \subset \widehat{\partial}\varphi(\bar{x}_0^N,\bar{x}_{k+1}^N) + 2\|\bar{x}_0^N - \bar{x}(a)\| I\!B^*\\ &+ \sum_{j=-N}^{-1} \Big[\int_{t_j}^{t_{j+1}} 2\|\bar{x}_j^N - \bar{x}(t)\|\,dt\Big] I\!B^* + h_N \sum_{j=0}^{k} \widehat{\partial} f(\bar{x}_j^N,\bar{x}_{j-N}^N,\bar{y}_j^N,t_j)\\ &+ \sum_{j=0}^{k} \Big[\int_{t_j}^{t_{j+1}} 2\|\bar{y}_j^N - \dot{\bar{x}}(t)\|\,dt\Big] I\!B^* + \varepsilon_N h_N I\!B^*,\end{aligned}$$

where the Fréchet subdifferential of the function f is considered with respect of its all but t variables, and where the classical relationship $\partial\|\cdot\|^2(x) \subset 2\|x\| I\!B^*$ is used together with the subdifferentiation formula under the integral sign in (5.3) well known from convex analysis. Substituting the latter relationships into (5.16) and adjusting ε_N if necessary, we arrive at

$$\begin{aligned}&0 \in \widehat{\partial}\varphi(\bar{x}_0^N,\bar{x}_{k+1}^N) + 2\|\bar{x}_0^N - \bar{x}(a)\| I\!B^* + \sum_{j=-N}^{-1} \Big[\int_{t_j}^{t_{j+1}} 2\|\bar{x}_j^N - \bar{x}(t)\|\,dt\Big] I\!B^*\\ &\quad + h_N \sum_{j=0}^{k} \widehat{\partial} f(\bar{x}_j^N,\bar{x}_{j-N}^N,\bar{y}_j^N,t_j) + \sum_{j=0}^{k} \Big[\int_{t_j}^{t_{j+1}} 2\|\bar{y}_j^N - \dot{\bar{x}}(t)\|\,dt\Big] I\!B^*\end{aligned} \tag{5.18}$$

$$+\sum_{j=0}^{k}\nabla g_j(\bar z^N)^* e_j^* + \sum_{j=-N}^{k+1}\widehat N(\bar z^N;\Theta_j)+\varepsilon_N h_N I\!B^*.$$

To elaborate the relationships in (5.18), let $z_j^* = (x^*_{-N,j},\ldots,x^*_{k+1,j},y^*_{0,j},\ldots,y^*_{k,j})$ and observe from the set structures in (5.6) that for any $z_j^*\in\widehat N(\bar z^N;\Theta_j)$, $j=-N,\ldots,-1$, *all but one* components of z_j^* are *zero* with the remaining one satisfying $x^*_{j,j}\in\widehat N(\bar x_j^N;C(t_j))$, $j=-N,\ldots,-1$. Similarly the relationships $z_j^*\in\widehat N(\bar z^N;\Theta_j)$ for $j=0,\ldots,k$ and $z^*_{k+1}\in\widehat N(\bar z^N;\Theta_{k+1})$ imply that

$$\begin{aligned}&(x^*_{j,j},x^*_{j-N,j},y^*_{j,j})\in\widehat N((\bar x_j^N,\bar x_{j-N}^N,\bar y_j^N);\operatorname{gph}F_j)\quad\text{for } j=0,\ldots,k,\\&(x^*_{0,k+1},x^*_{k+1,k+1})\in\widehat N((\bar x_0^N,\bar x_{k+1}^N);\Omega_N)\end{aligned}\tag{5.19}$$

with all the other components of z_j^*, $j=0,\ldots,k+1$, equal to zero. Combining these relationships with (5.17) and (5.18) and using the notation

$$(u_0^N,u_{k+1}^N)\in\widehat\partial\varphi(\bar x_0^N,\bar x_{k+1}^N),\quad (v_j^N,\kappa_{j-N}^N,w_j^N)\in\widehat\partial f(\bar x_j^N,\bar x_{j-N}^N,\bar y_j^N,t_j),$$

and (5.9) for (θ_j^N,σ_j^N) with $\bar y_j^N=(\bar x_{j+1}^N-\bar x_j^N)/h_N$ due to $g(\bar z^N)=0$ in (5.5), we arrive at

$$\begin{cases}-x^*_{j,j}-x^*_{j,j+N}\in h_N\kappa_j^N+\sigma_j^N I\!B^*+\varepsilon_N h_N I\!B^*,\quad j=-N,\ldots,-1,\\ -x^*_{j,j}-x^*_{j,j+N}\in h_N\kappa_j^N+h_N v_j^N+e^*_{j-1}-e_j^*+\varepsilon_N h_N I\!B^*,\\ \qquad\qquad\qquad\qquad\qquad\qquad j=1,\ldots,k-N,\\ -x^*_{j,j}\in h_N v_j^N+e^*_{j-1}-e^*_j+\varepsilon_N h_N I\!B^*,\quad j=k-N+1,\ldots,k,\\ -y^*_{j,j}\in h_N w_j^N+\theta_j^N I\!B^*-h_N e_j^*+\varepsilon_N h_N I\!B^*,\quad j=0,\ldots,k,\\ -x^*_{k+1,k+1}\in u^N_{k+1}+e_k^*+\varepsilon_N h_N I\!B^*,\\ -x^*_{0,0}-x^*_{0,k+1}\in u_0^N+h_N\kappa_0^N+2\|\bar x_0^N-\bar x(a)\|I\!B^*\\ \qquad +h_N v_0^N-e_0^*+\varepsilon_N h_N I\!B^*,\end{cases}\tag{5.20}$$

where $(x^*_{j,j},x^*_{j-N,j},y^*_{j,j})$ and $(x^*_{0,k+1},x^*_{k+1,k+1})$ satisfy (5.19). Further, let

$$\begin{cases}\widetilde p_j^N:=e^*_{j-1}\ \text{ for }\ j=1,\ldots,k+1,\\ \widetilde q_j^N:=\kappa_j^N+\dfrac{x^*_{j,j+N}}{h_N}\ \text{ for }\ j=-N,\ldots,k-N,\\ \widetilde q_j^N:=0\ \text{ for }\ j=k-N+1,\ldots,k+1\end{cases}\tag{5.21}$$

and define the the *adjoint discrete trajectories* (p_j^N,q_j^N) by

$$\begin{cases}q^N_{k+1}:=0,\quad q_j^N:=q^N_{j+1}-\widetilde q_j^N h_N\ \text{ for }\ j=-N,\ldots,k+1,\\ p_0^N:=u_0^N+x^*_{0,k+1}-q_0^N,\\ p_j^N:=\widetilde p_j^N-q_j^N h_N\ \text{ for }\ j=1,\ldots,k+1.\end{cases}\tag{5.22}$$

It is easy to check that $q_j^N=0$ for $j=k-N+1,\ldots,k+1$. Combining finally the relationships and notation (5.19)–(5.22), we get the optimality conditions

(5.10)–(5.13) of the theorem with $\lambda^N = 1$ along an arbitrarily chosen sequence $\varepsilon_N \downarrow 0$ as $N \to \infty$. This completes the proof in Case 1.

Case 2. It remains to consider the situation when the mapping g from (5.5) is *not metrically regular* at $\bar{z}^N$ relative to the set Θ. In this case the *restriction* of g on Θ defined by

$$g_\Theta(z) := \begin{cases} g(z) & \text{if } z \in \Theta, \\ \emptyset & \text{otherwise} \end{cases} \tag{5.23}$$

is not metrically regular around $\bar{z}^N$ in the standard sense. Picking again an arbitrary sequence $\varepsilon_N \downarrow 0$ as $N \to \infty$ and using the *characterization* of the metric regularity property from Theorem 4.5 in Mordukhovich (2006a), for any fixed $N \in I\!N$ we find $z \in \bar{z}^N + \varepsilon_N I\!B$ and $e^* = (e_0^*, \ldots, e_k^*) \in (X^*)^k$ satisfying the relationships

$$\|e^*\| > 1 \ \text{ and } \ 0 \in \widehat{D}^* g_\Theta(z)(e^*). \tag{5.24}$$

Taking now into account the structure of the mapping g_Θ in (5.23) and applying the *coderivative sum rule* for (4.5) from Theorem 1.62 in Mordukhovich (2006a) and then the afore-mentioned *intersection rule* for Fréchet normals to Θ in (5.9), we get

$$0 \in \sum_{j=0}^{k} \nabla g_j(z)^* e_j^* + \sum_{j=-N}^{k+1} \widehat{N}(z_j; \Theta_j) + \varepsilon_N h_N I\!B^*$$

with some $z_j \in \Theta_j \cap (z + \varepsilon_N I\!B)$. Thus there are $z_j^* \in \widehat{N}(\bar{z}^N; \Theta_j)$, $j = -N, \ldots, k+1$, such that

$$-\sum_{j=-N}^{k+1} z_j^* \in \sum_{j=0}^{k} \nabla g_j(\bar{z}^N)^* e^* + \varepsilon_N h_N I\!B^* \tag{5.25}$$

It follows from (5.25), (5.17), and the corresponding arguments in Case 1 that there are $(x_{j,j}^*, x_{j-N,j}^*, y_{j,j}^*)$ and $(x_{0,k+1}^*, x_{k+1,k+1}^*)$ satisfying (5.19), for which

$$\begin{cases} -x_{j,j}^* - x_{j,j+N}^* \in \varepsilon_N h_N I\!B^*, \ j = -N, \ldots, -1, \\ -x_{j,j}^* - x_{j,j+N}^* \in e_{j-1}^* - e_j^* + \varepsilon_N h_N I\!B^*, \ j = 1, \ldots, k-N, \\ -x_{j,j}^* \in e_{j-1}^* - e_j^* + \varepsilon_N h_N I\!B^*, \ j = k-N+1, \ldots, k, \\ -y_{j,j}^* \in h_N e_j^* + \varepsilon_N h_N I\!B^*, \ j = 0, \ldots, k, \\ -x_{k+1,k+1}^* \in e_k^* + \varepsilon_N h_N I\!B^*, \\ -x_{0,0}^* - x_{0,k+1}^* \in -e_0^* + \varepsilon_N h_N I\!B^*. \end{cases}$$

Defining further the *adjoint discrete trajectories* p_j^N for $j = 0, \ldots, k+1$ and q_j^N for $j = -N, \ldots, k+1$ in the same way as in Case 1, we justify by similar arguments the validity of the *approximate Euler-Lagrange inclusion* (5.11), the *approximate tail conditions* (5.12), and the *approximate transversality inclusion* (5.13) with $\lambda^N = 0$. Let us now verify that the local *Lipschitz continuity* of F assumed in (H2) implies the fulfillment of the *nontriviality condition* (5.10).

First we show that there exist two positive numbers α_1 and α_2 independent of N such that

$$\|p_j^N\| \le \alpha_1 \|p_{k+1}^N\| + \alpha_2 \varepsilon_N, \quad j = 0, \ldots, k. \tag{5.26}$$

Observe that the approximate Euler-Lagrange inclusion (5.11) with $\lambda^N = 0$ can be equivalently written in terms of the *coderivative* (4.5) as

$$\Big(\frac{p_{j+1}^N - p_j^N}{h_N}, \frac{q_{j-N+1}^N - q_{j-N}^N}{h_N}\Big) \in \widehat{D}^* F(\bar{x}_j^N, \bar{x}_{j-N}^N, \bar{y}_j^N)(-p_{j+1}^N - q_{j+1}^N) + \varepsilon_N I\!B^*.$$

Then, using the neighborhood *characterization* of the local *Lipschitzian property* from Theorem 4.7 in Mordukhovich (2006a), we get that

$$\Big\|\Big(\frac{p_{j+1}^N - p_j^N}{h_N}, \frac{q_{j-N+1}^N - q_{j-N}^N}{h_N}\Big)\Big\| \le L_F \|p_{j+1}^N + q_{j+1}^N\| + \varepsilon_N, \quad j = 0, \ldots, k, \tag{5.27}$$

where L_F is the Lipschitz constant of F from (2.2) in (H2). Noting that $q_{j+1}^N = 0$ for $j = k - N, \ldots, k$ by (5.12), we have for these indices that

$$\begin{aligned}
&\|(p_j^N, q_{j-N}^N)\| \le L_F h_N \|p_{j+1}^N\| + \|(p_{j+1}^N, q_{j-N+1}^N)\| + h_N \varepsilon_N \\
&\le (L_F h_N + 1)\|(p_{j+1}^N, q_{j-N+1}^N)\| + h_N \varepsilon_N \\
&\le (L_F h_N + 1)^2 \|(p_{j+2}^N, q_{j-N+2}^N)\| + (L_F h_N + 1) h_N \varepsilon_N + h_N \varepsilon_N \\
&\le \ldots \\
&\le (L_F h_N + 1)^{k+1-j}[\|p_{k+1}^N\| + \varepsilon_N / L_F] \\
&\le (L_F h_N + 1)^{N+1}[\|p_{k+1}^N\| + \varepsilon_N / L_F] \le e^{L_F \Delta}[\|p_{k+1}^N\| + \varepsilon_N / L_F].
\end{aligned}$$

For the indices $j = k - 2N, \ldots, k - N - 1$ we get from (5.27) and the estimates above that

$$\begin{aligned}
&\|(p_j^N, q_{j-N}^N)\| \le L_F h_N \|p_{j+1}^N + q_{j+1}^N\| + \|(p_{j+1}^N, q_{j-N+1}^N)\| + h_N \varepsilon_N \\
&\le (L_F h_N + 1)\|(p_{j+1}^N, q_{j-N+1}^N)\| \\
&\quad + L_F h_N (e^{L_F \Delta}\|p_{k+1}^N\| + e^{L_F \Delta} \varepsilon_N / L_F) + h_N \varepsilon_N + (L_F h_N + 1) h_N \varepsilon_N \\
&\quad + (L_F h_N + 1) L_F h_N (e^{L_F \Delta}\|p_{k+1}^N\| + e^{L_F \Delta} \varepsilon_N / L_F) \\
&\le \ldots \\
&\le (L_F h_N + 1)^{k+1-j}[\|(p_{k+1}^N\| + e^{L_F \Delta}\|p_{k+1}^N\| + e^{L_F \Delta} \varepsilon_N / L_F + \varepsilon_N / L_F] \\
&\le e^{L_F \Delta}(1 + e^{L_F \Delta})[\|p_{k+1}^N\| + \varepsilon_N / L_F].
\end{aligned}$$

After repeating the above process finitely many times we arrive at the desired estimate (5.26).

To conclude now the proof of the nontriviality condition (5.10) along with (5.11)–(5.13) and $\lambda^N = 0$, suppose the *opposite* and then, taking a sequence $\gamma_m \downarrow 0$ as $m \to \infty$, choose numbers $N_m \in I\!N$ and $\widetilde{\varepsilon}_m := \varepsilon_{N_m} > 0$ such that

$$k_m := [1/\gamma_m], \quad \widetilde{\varepsilon}_m \le \gamma_m^2, \quad \text{and} \quad \|p_{k_m+1}^{N_m}\| \le \gamma_m^2 \quad \text{as} \quad m \in I\!N,$$

where k_m is computed by (2.4) for N_m, and where $[\cdot]$ stands for the greatest integer less than or equal to the given real number. Then by (5.26) we have

$$\sum_{j=1}^{k_m+1} \|p_j^{N_m}\| \le \alpha_1(k_m+1)\gamma_m^2 + \alpha_2\widetilde{\varepsilon}_m(k_m+1) \le 2(\alpha_1+\alpha_2)\gamma_m \downarrow 0 \quad \text{as} \quad m \to \infty,$$

which *contradicts* the negation of metric regularity (5.24) imposed in Case 2 and thus completes the proof of the theorem. ■

6. Euler-Lagrange conditions for delay-differential inclusions

In this section we derive *necessary optimality conditions* for the given optimal solution $\bar{x}(\cdot)$ to the original Bolza problem (P). The proof is based on the passing to the limit from the necessary optimality conditions for the discrete approximation problems (P_N) obtained in Section 5. We keep assumptions (H1–(H3) and (H6), but instead of (H4) and (H5) impose their following modifications:

(H4') φ is Lipschitz continuous on $U \times U$; $\Omega = \Omega_a \times \Omega_b \subset X \times X$, where Ω_a is compact around $\bar{x}(a)$ while Ω_b is closed around $\bar{x}(b)$.

(H5') The integrand $f(x, y, v, \cdot)$ is continuous for a.e. $t \in [a, b]$ and bounded uniformly with respect to $(x, y, v) \in U \times (M_C I\!B) \times (M_F I\!B)$; furthermore, there are numbers $\mu > 0$ and $L_f \ge 0$ such that $f(\cdot,\cdot,\cdot,t)$ is Lipschitz continuous on the set $A_\mu(t)$ from (H5) with constraint L_f uniformly in $t \in [a, b]$.

The next theorem establishes necessary optimality conditions in the *extended Euler-Lagrange form* for the given optimal solution to the original problem (P) in terms of the limiting normals and subgradients of Section 4 for the initial data of (P) computed with respect to all but time variables along the reference optimal solution. Note that the optimality conditions obtained in the general case of *geometric* endpoint constraints in *infinite-dimensional* state spaces require the *sequential normal compactness* assumption imposed on Ω_b at the optimal endpoint $\bar{x}(b)$.

THEOREM 4 **(extended Euler-Lagrange conditions for delay-differential inclusions).** *Let $\bar{x}(\cdot)$ be an optimal solution to (P) under hypotheses* (H1)–(H3), (H4'), (H5'), *and* (H6). *Assume in addition that both spaces X and X^**

are Asplund, that Ω_b is SNC at $\bar{x}(b)$, and that (P) is stable with respect to relaxation. Then there exist a number $\lambda \geq 0$ and two absolutely continuous adjoint arcs $p\colon [a,b] \to X^*$ and $q\colon [a-\Delta, b] \to X^*$ such that the following conditions hold:

—the extended Euler-Lagrange inclusion

$$\begin{aligned}(\dot{p}(t), \dot{q}(t-\Delta)) \in \operatorname{clco}\big\{&(u,w) \mid (u,w,p(t)+q(t)) \\ &\in \lambda\partial_+ f(\bar{x}(t), \bar{x}(t-\Delta), \dot{\bar{x}}(t), t) \\ &+ N_+((\bar{x}(t), \bar{x}(t-\Delta), \dot{\bar{x}}(t)); \operatorname{gph} F(\cdot,\cdot,t))\big\} \quad \text{a.e. } t \in [a,b],\end{aligned} \tag{6.1}$$

*where the norm-closure operation "*cl*" can be omitted when the state space X is reflexive;*

—the optimal tail conditions

$$\begin{cases} \langle \dot{q}(t), \bar{x}(t)\rangle = \min\limits_{c \in C(t)} \langle \dot{q}(t), c\rangle \quad \text{a.e. } t \in [a-\Delta, a), \\ q(t) = 0, \quad t \in [b-\Delta, b]; \end{cases} \tag{6.2}$$

—the transversality inclusion

$$(p(a)+q(a), -p(b)) \in \lambda\partial\varphi(\bar{x}(a), \bar{x}(b)) + N(\bar{x}(a); \Omega_a) \times N(\bar{x}(b); \Omega_b); \tag{6.3}$$

— the nontriviality condition

$$\lambda + \|p(b)\| > 0. \tag{6.4}$$

Proof. We derive the optimality conditions of the theorem by passing to the limit in the necessary optimality conditions obtained in Theorem 3 and using the strong convergence of discrete approximations established in Theorem 2. We actually need more: to justify a suitable convergence of *adjoint/dual* elements in the necessary optimality conditions for discrete approximations. It is done in what follows by employing the afore-mentioned coderivative characterization of Lipschitz continuity, robustness of our limiting generalized differential constructions, and the imposed SNC property of the endpoint constraint set together with appropriate facts of functional analysis.

Recall again that the Asplund property of both spaces X and X^* ensures the Radon-Nikodým property of these spaces. This implies, in particular, that the absolute continuity of the primal and adjoint arcs in the setting of the theorem is equivalent to the fulfillment of the Newton-Leibniz formula (1.4) for these arcs. Note also that the assumptions made in this theorem ensure the validity of all the assumptions made in both Theorem 2 and Theorem 3.

Employing the necessary optimality conditions for (P_N) obtained in Theorem 3, we find sequences of numbers $\lambda^N \geq 0$ and adjoint discrete trajectories p_j^N and q_{j-N}^N satisfying inclusions (5.10)-(5.13) with some $\varepsilon_N \downarrow 0$ as $N \to \infty$. Observe that without loss of generality the nontriviality condition (5.10) can be equivalently written as

$$\lambda^N + \|p_{k+1}^N\| = 1 \;\text{ for all }\; N \in I\!N, \tag{6.5}$$

since the number $\gamma > 0$ in (5.10) is independent of N.

Suppose, without loss of generality, that $\lambda^N \to \lambda \geq 0$ as $N \to \infty$. As above, the notation $\bar{x}^N(t)$, $p^N(t)$, and $q^N(t-\Delta)$ indicates the piecewise linear extensions of the discrete arcs to the corresponding continuous-time intervals with their piecewise constant derivatives $\dot{\bar{x}}^N(t)$, $\dot{p}^N(t)$, and $\dot{q}^N(t-\Delta)$. Based on (5.9), define their piecewise constant extensions

$$\theta^N(t) := \frac{\theta_j^N}{h_N} a_j^N \quad \text{for} \quad t \in [t_j, t_{j+1}), \quad j = 0, \ldots, k,$$
$$\sigma^N(t) := \frac{\sigma_j^N}{h_N} b_j^N \quad \text{for} \quad t \in [t_j, t_{j+1}), \quad j = -N, \ldots, -1,$$

and conclude from the *strong convergence* results of Theorem 1 that

$$\begin{aligned}
\int_a^b \|\theta^N(t)\|\,dt &= \sum_{j=0}^{k} \|\theta_j^N\| \leq 2\sum_{j=0}^{k} \int_{t_j}^{t_{j+1}} \Big\| \frac{\bar{x}_{j+1}^N - \bar{x}_j^N}{h_N} - \dot{\bar{x}}(t) \Big\|\,dt \\
&= 2\int_a^b \|\dot{\bar{x}}^N(t) - \dot{\bar{x}}(t)\|\,dt \to 0, \\
\int_{a-\Delta}^{a} \|\sigma^N(t)\|\,dt &= \sum_{j=-N}^{-1} \|\sigma_j^N\| \leq 2\sum_{j=-N}^{-1} \int_{t_j}^{t_{j+1}} \|\bar{x}_j^N - \bar{x}(t)\|\,dt \\
&= 2\int_{a-\Delta}^{a} \|\bar{x}^N(t) - \bar{x}(t)\|\,dt \to 0
\end{aligned}$$

as $N \to \infty$. Since the strong L^1 convergence of $\{\theta^N(\cdot)\}$ and $\{\sigma^N(\cdot)\}$, established above, implies the a.e. convergence of their subsequences, we suppose without loss of generality that

$$\theta^N(t) \to 0 \;\text{ a.e. }\; t \in [a,b],\; \sigma^N(t) \to 0 \;\text{ a.e. }\; t \in [a-\Delta, a] \;\text{ as }\; N \to \infty. \tag{6.6}$$

Further, let us estimate $(p^N(t), q^N(t-\Delta))$ for large N. It follows from the approximate Euler-Lagrange condition (5.11) that for all $j = 0, \ldots, k$ we have the inclusions

$$\begin{aligned}
&\Big(\frac{p_{j+1}^N - p_j^N}{h_N} - \lambda^N v_j^N, \frac{q_{j-N+1}^N - q_{j-N}^N}{h_N} - \lambda^N \kappa_{j-N}^N, \\
&\quad -\frac{\lambda^N \theta_j^N}{h_N} a_j^N + p_{j+1}^N + q_{j+1}^N - \lambda^N \omega_j^N\Big) \\
&\in \widehat{N}((\bar{x}_j^N, \bar{x}_{j-N}^N, \bar{y}_j^N); \operatorname{gph} F_j) + \varepsilon_N I\!B^*
\end{aligned}$$

with some $(v_j^N, \kappa_{j-N}^N, \omega_j^N) \in \widehat{\partial} f(\bar{x}_j^N, \bar{x}_{j-N}^N, \bar{y}_j^N, t_j)$ and $a_j^N \in I\!B^*$. This implies by (4.5) that

$$\begin{aligned}&\Big(\frac{p_{j+1}^N-p_j^N}{h_N}-\lambda^N v_j^N,\frac{q_{j-N+1}^N-q_{j-N}^N}{h_N}-\lambda^N\kappa_{j-N}^N\Big)\\&\in\widehat{D}^*F_j(\bar{x}_j^N,\bar{x}_{j-N}^N,\bar{y}_j^N)\Big(\lambda^N\omega_j^N+\frac{\lambda^N\theta_j^N}{h_N}a_j^N-p_{j+1}^N-q_{j+1}^N\Big)+\varepsilon_N I\!B^*\end{aligned}$$

for these indices j, which gives by the *coderivative condition* for Lipschitzian stability taken from Theorem 1.43 in Mordukhovich (2006a) that

$$\begin{aligned}&\Big\|\Big(\frac{p_{j+1}^N-p_j^N}{h_N}-\lambda^N v_j^N,\frac{q_{j-N+1}^N-q_{j-N}^N}{h_N}-\lambda^N\kappa_{j-N}^N\Big)\Big\|\\&\le L_F\Big\|\lambda^N\omega_j^N+\frac{\lambda^N\theta_j^N}{h_N}a_j^N-p_{j+1}^N-q_{j+1}^N\Big\|+\varepsilon_N,\quad j=0,\dots,k.\end{aligned}\tag{6.7}$$

The subdifferential specification of the latter result for the case of locally Lipschitzian functions ensures the estimates $\|(v_j^N,\kappa_{j-N}^N,\omega_j^N)\|\le L_f$ for $j=0,\dots,k$, which implies by (6.7) and the approximate tail conditions in (5.12) that

$$\begin{aligned}&\|(p_j^N,q_{j-N}^N)\|\le L_F\|\lambda^N\theta_j^N\|+L_F\lambda^N h_N\|\omega_j^N\|+L_F h_N\|p_{j+1}^N+q_{j+1}^N\|\\&\quad+\|(p_{j+1}^N,q_{j-N+1}^N)\|+\lambda^N h_N\|(v_j^N,\kappa_{j-N}^N)\|+h_N\varepsilon_N\\&\le L_F\|\theta_j^N\|+(L_F+1)h_N L_f+(L_F h_N+1)\|(p_{j+1}^N,q_{j-N+1}^N)\|+h_N\varepsilon_N\\&\le L_F\|\theta_j^N\|+(L_F h_N+1)L_F\|\theta_{j+1}^N\|\\&\quad+(L_F+1)h_N L_f+(L_F h_N+1)(L_F+1)h_N L_f\\&\quad+(L_F h_N+1)^2\|(p_{j+2}^N,q_{j-N+2}^N)\|+(L_F h_N+1)h_N\varepsilon_N+h_N\varepsilon_N\le\dots\\&\le\exp[L_F(b-a)](1+L_f(L_F+1)/L_F+L_F\nu_N)\\&\quad+[(L_F h_N+1)^N-1]\varepsilon_N/L_F\end{aligned}$$

for $j=k-N,\dots,k$, where

$$\nu_N:=\int_a^b\|\dot{\bar{x}}(t)-\dot{\bar{x}}^N(t)\|\,dt\to 0\ \text{ as }\ N\to\infty$$

by Theorem 2, and where $[(L_F h_N+1)^N-1]\varepsilon_N/L_F\to 0$ by $\varepsilon_N\to 0$ as $N\to\infty$. The latter implies the *uniform boundedness* of the sequence $\{(p_j^N,q_{j-N}^N)|\ j=k-N,\dots,k\}$ with respect to $N\in I\!N$, i.e., there is $M_1>0$ independent of N such that

$$\|(p_j^N,q_{j-N}^N)\|\le M_1\ \text{ for all }\ j=k-N,\dots,k\ \text{ and }\ N\in I\!N.\tag{6.8}$$

Thus, the the piecewise linear extensions $p^N(t)$ and $q^N(t-\Delta)$ are uniformly bounded on $[b-\Delta,b]$.

For $j = k-2N, \ldots, k-N-1$, it follows from (6.7) and (6.8) that

$$\begin{aligned}
&\|(p_j^N, q_{j-N}^N)\| \le L_F\|\lambda^N \theta_j^N\| + L_F\lambda^N h_N\|\omega_j^N\| + L_F h_N\|p_{j+1}^N + q_{j+1}^N\| \\
&\quad + \|(p_{j+1}^N, q_{j-N+1}^N)\| + \lambda^N h_N\|(v_j^N, \kappa_{j-N}^N)\| + h_N\varepsilon_N \\
&\le L_F\|\theta_j^N\| + (L_F h_N + h_N)L_f + L_F h_N M_1 + L_F h_N\|p_{j+1}^N\| \\
&\quad + \|(p_{j+1}^N, q_{j-N+1}^N)\| + h_N\varepsilon_N \\
&\le L_F\|\theta_j^N\| + (L_F+1)h_N L_f + L_F h_N M_1 \\
&\quad + (L_F h_N + 1)\|(p_{j+1}^N, q_{j-N+1}^N)\| + h_N\varepsilon_N,
\end{aligned}$$

which implies the uniform boundedness of the sequence $\{(p_j^N, q_{j-N}^N)|\ j = k - 2N, \ldots, k-N+1\}$ as $N \in I\!N$ and hence the uniform boundedness of $\{(p^N(t), q^N(t-\Delta))\}$ on the interval $[b-2\Delta, b-\Delta]$. Repeating the above procedure, we conclude that the sequence $\{(p^N(t), q^N(t-\Delta))\}$ is *uniformly bounded* on the *whole interval* $[a,b]$.

To estimate $(\dot p^N(t), \dot q^N(t-\Delta))$, we have from (6.7) that

$$\begin{aligned}
&\|(\dot p^N(t), \dot q^N(t-\Delta))\| = \Big\|\Big(\frac{p_{j+1}^N - p_j^N}{h_N}, \frac{q_{j-N+1}^N - q_{j-N}^N}{h_N}\Big)\Big\| + \varepsilon_N \\
&\le L_F\Big\|\lambda^N\omega_j^N + \frac{\lambda^N\theta_j^N}{h_N}a_j^N - p_{j+1}^N - q_{j+1}^N\Big\| + \lambda^N\|(v_j^N, \kappa_{j-N}^N)\| + \varepsilon_N \\
&\le (L_F+1)L_f + L_F(\|\theta^N(t)\| + \|p_{j+1}^N\| + \|q_{j+1}^N\|) + \varepsilon_N,\ t \in [t_j, t_{j+1}),
\end{aligned} \tag{6.9}$$

for all $j = 0, \ldots, k$ and $N \in I\!N$. Taking into account (6.6) and the uniform boundedness of $\{(p^N(t), q^N(t-\Delta))\}$ as well as the RNP of both X and X^*, we apply the afore-mentioned Dunford theorem on the (sequential) *weak compactness* in $L^1([a,b];X^*)$ and conclude with no loss of generality that both sequences $\{\dot p^N(t)\}$ and $\{\dot q^N(t-\Delta)\}$ *weakly converge* in $L^1([a,b];X^*)$. Furthermore, by $\|p^N(b)\| \le 1$ as $N \in I\!N$ due to (6.5) and the Asplund property of X we have that $\{p^N(b)\}$ is *sequentially weak* compact* in X^*. Arguing now as in the proof of Theorem 2 by using the Newton-Leibniz formula for $p^N(\cdot)$ and the weak continuity of the Bochner integral as a linear operator from $L^1([a,b];X^*)$ to X^*, we get an absolutely continuous arc $p\colon [a,b] \to X^*$ such that

$$\begin{cases} p^N(t) \to p(t) \ \text{ weak}^* \text{ in } \ X^* \ \text{ for all } \ t \in [a,b], \\ \dot p^N(\cdot) \to \dot p(\cdot) \ \text{ weakly in } \ L^1([a,b];X^*) \ \text{ as } \ N \to \infty. \end{cases} \tag{6.10}$$

Similarly, by taking into account the second tail condition in (5.12), we find an absolutely continuous arc $q : [a-\Delta, b] \to X^*$ such that $q(t)$ satisfies the *second tail condition* on $[b-\Delta, b]$ in (6.2) and

$$\begin{cases} q^N(t-\Delta) \to q(t-\Delta) \ \text{ weak}^* \text{ in } X^* \ \text{ for all } \ t \in [a,b], \\ \dot q^N(\cdot-\Delta) \to \dot q(\cdot-\Delta) \ \text{ weakly in } \ L^1([a,b];X^*) \ \text{ as } \ N \to \infty. \end{cases} \tag{6.11}$$

The *first tail condition* on $[a-\Delta, a]$ in (6.2) follows by passing to the limit in the corresponding one from (5.12), taking into account the convergence in (6.11) and of $\sigma^N(\cdot)$ in (6.6) and the specific structure of the normal cone to *convex* sets given in (4.3).

To prove the *extended Euler-Lagrange inclusion* (6.1) by passing to the limit in the approximate one (5.11), we rewrite the latter as

$$\begin{aligned}&(\dot p^N(t),\dot q^N(t-\Delta))\in\big\{(u,v)\mid\big(u,v,p^N(t_{j+1})+q^N(t_{j+1})-\lambda^N\theta_j^N a_j^N/h_N\big)\\&\in\lambda^N\widehat\partial f(\bar x(t_j),\bar x(t_j-\Delta),\dot{\bar x}^N(t_j),t_j)\\&+\widehat N\big((\bar x^N(t_j),\bar x^N(t_j-\Delta),\dot{\bar x}^N(t));\operatorname{gph}F(\cdot,\cdot,t_j)\big)\big\}+\varepsilon_N I\!B^*\end{aligned}\tag{6.12}$$

for $t\in[t_j,t_{j+1})$, $j=0,\dots,k$, and $N\in I\!N$. Observe that the weak convergence in $L^1([a,b];X^*)$ of the derivatives $\dot p^N(\cdot)$ and $\dot q^N(\cdot-\Delta)$ from (6.10) and (6.11) implies by the classical Mazur theorem the strong convergence in $L^1([a,b];X^*)$ of their *convex combinations* and hence the a.e. *pointwise* convergence of (some subsequences of) these combinations on $[a,b]$. Using this, the weak* pointwise convergence in X^* of $\{(p^N(t),q^N(t-\Delta))\}$ from (6.10) and (6.11), the pointwise convergence of $\{\theta^N(t)\}$ from (6.6), the strong convergence of $\{\dot{\bar x}^N(t)\}$ from Theorem 2, and the constructions of extended limiting normals and subgradients from (4.8) and (4.9), we pass to the limit in (6.12) as $N\to\infty$ and arrive at the extended Euler-Lagrange inclusion (6.1).

If X is *reflexive*, the *closure operation* in (6.1) *can be omitted.* Indeed, in the reflexive case weak and weak* topology agree and, furthermore, every bounded and convex set is weakly compact in X^*, being therefore automatically *closed* in the *norm topology* of X^* due the afore-mentioned Mazur theorem. Hence, the arguments above allow us to drop the closure operation in the limiting convexification procedure due to the derivative estimates in (6.9).

To derive the *transversality inclusion* in (6.3), we pass to the limit in the approximate one from (5.13) as $N\to\infty$. Since $\Omega_N=\Omega+\eta I\!B$ in (5.13) with $\eta_N\to 0$ as $N\to\infty$ by Theorem 1, we first employ the *sum rule* for Fréchet normals from Theorem 3.7(i) in Mordukhovich (2006a) and then pass to limiting normals and subgradients in (6.3) by using the weak* convergence of $\{p^N(a)\}$ and of $\{p^N(b)\}$ in X^* and the simple formula for basic normals to the Cartesian product of sets.

To complete the proof of the theorem, it remains to verify the *nontriviality condition* (6.4) under the *SNC* assumption on Ω_b at $\bar x(b)$. Suppose, on the contrary, that $\lambda=0$ and $p(b)=0$ for the limiting elements in the above procedure. Without loss of generality, assume that $\lambda^N=0$ for all $N\in I\!N$. It follows from the arguments above that $p^N(b)\xrightarrow{w^*}0$ as $N\to\infty$ in this case. By the approximate transversality condition (5.13) with $\lambda^N=0$ we have that

$$-p^N(b)\in\widehat N(\bar x^N(b);\Omega_b+\eta_N I\!B)+\varepsilon_N I\!B^*.$$

Applying then the afore-mentioned sum rule for Fréchet normals to the latter inclusion and taking into account its structure, we find a sequence $\{\widetilde p^N\}\subset X^*$

satisfying

$$-\widetilde{p}^N \in \widehat{N}(\bar{x}^N(b);\Omega_b) \quad \text{and} \quad \|\widetilde{p}^N - p^N(b)\| \to 0 \quad \text{as} \quad N \to \infty. \tag{6.13}$$

Thus, $\widetilde{p}^N \stackrel{w^*}{\to} 0$ and, by the assumed SNC property of Ω_b at $\bar{x}(b)$, we get from the first relationship in (6.13) that $\|\widetilde{p}^N\| \to 0$ as $N \to \infty$. This implies by the second relationship in (6.13) that $\|p^N(b)\| = \|p^N(t_{k+1})\| \to 0$ as $N \to \infty$, which clearly contradicts the nontriviality condition (6.5) for discrete approximations. This completes the proof of the theorem. ■

7. Delay systems with functional endpoint constraints

In the last section of the paper we consider a version of the generalized Bolza problem (P) formulated in Section 1, where endpoint constraints of the geometric type (1.3) are replaced by their *functional* counterpart given by *finitely many* equalities and inequalities with *Lipschitz continuous* functions. Let us denote this problem by (P_0) and describe as follows, where for simplicity we confine ourselves to the case of delay-differential inclusions with *fixed* left endpoints:

$$\text{minimize} \quad J[x] := \varphi_0(x(b)) + \int_a^b f(x(t), x(t-\Delta), \dot{x}(t), t)\, dt$$

over feasible arcs $x\colon [a-\Delta, b] \to X$ as for (P) in Section 1 with $\Delta > 0$, subject to

$$\begin{aligned}
&\dot{x}(t) \in F(x(t), x(t-\Delta), t) \quad a.e.\ t \in [a,b], \quad x(a) = x_0 \in X,\\
&x(t) \in C(t) \quad a.e.\ t \in [a-\Delta, a),\\
&\varphi_i(x(b)) \le 0, \quad i = 1, \ldots, m,\\
&\varphi_i(x(b)) = 0, \quad i = m+1, \ldots, m+r.
\end{aligned}$$

Given an *optimal solution* $\bar{x}(\cdot)$ to (P_0), we keep assumptions (H1)–(H3), (H5'), and (H6) while replace (H4) and (H4') by the following:

(H4") The cost function φ_0 and all the endpoint constraint functions φ_i, $i = 1, \ldots, m+r$, are locally Lipschitzian around $\bar{x}(b)$.

The next theorem provides necessary optimality conditions for the given optimal solution $\bar{x}(\cdot)$ to (P_0) in the *extended Euler-Lagrange form* with the transversality inclusion expressed via the basic subgradients of the endpoint functions. Observe the *different subdifferential treatments* therein of the *equality* constraints versus those for the *cost/inequality* ones given by *nonsmooth* functions and also the fact that *all multipliers* λ_i, $i = 0, \ldots, m+r$, are *nonnegative*. The main distinction between the results obtained for (P_0) and those in Theorem 4 for (P) is that we now do *not impose* the *SNC* assumption on the endpoint constraints. This is a remarkable specific feature of the constraints described by *finitely many* equalities and inequalities with *Lipschitzian* functions.

Theorem 5 **(extended Euler-Lagrange conditions for delay-differential inclusions with functional endpoint constraints).** *Let $\bar{x}(\cdot)$ be an optimal solution to problem (P_0) under hypotheses* (H1)–(H3), (H4"), (H5'), *and* (H6). *Assume in addition that both spaces X and X^* are Asplund and that problem (P_0) is stable with respect to relaxation. Then there are multipliers $(\lambda_0, \ldots, \lambda_{m+r}) \in I\!R^{m+r+1}$ and absolutely continuous dual arcs $p\colon [a,b] \to X^*$ and $q\colon [a-\Delta, b] \to X^*$ satisfying the following relationships:*
—the sign and nontriviality conditions

$$\lambda_i \geq 0 \ \ \text{for all} \ \ i = 0, \ldots, m+r, \quad \text{and} \ \sum_{i=0}^{m+r} \lambda_i \neq 0;$$

—the complementary slackness conditions

$$\lambda_i \varphi_i(\bar{x}(b)) = 0 \ \ \text{for} \ \ i = 1, \ldots, m;$$

—the extended Euler-Lagrange inclusion

$$(\dot{p}(t), \dot{q}(t-\Delta)) \in \operatorname{clco}\big\{(u,w) \mid (u, w, p(t)+q(t)) \in \lambda \partial_+ f(\bar{x}(t), \bar{x}(t-\Delta), \dot{\bar{x}}(t), t) + N_+((\bar{x}(t), \bar{x}(t-\Delta), \dot{\bar{x}}(t)); \operatorname{gph} F(\cdot,\cdot,t))\big\} \quad \text{a.e. } t \in [a,b],$$

where the norm-closure operation can be omitted when the state space X is reflexive;
—the optimal tail conditions

$$\begin{cases} \langle \dot{q}(t), \bar{x}(t) \rangle = \min\limits_{c \in C(t)} \langle \dot{q}(t), c \rangle & \text{a.e. } t \in [a-\Delta, a), \\ q(t) = 0, \quad t \in [b-\Delta, b]; \end{cases}$$

—the transversality inclusion

$$-p(b) \in \sum_{i=0}^{m} \lambda_i \partial \varphi_i(\bar{x}(b)) + \sum_{i=m+1}^{m+r} \lambda_i \big[\partial \varphi_i(\bar{x}(b)) \cup \partial(-\varphi_i)(\bar{x}(b))\big].$$

Proof. (Sketch of the proof) Let us discuss the following *two schemes* to justify the formulated optimality conditions. The *first one* goes in the direction developed by Mordukhovich (2007) for the case of nondelayed autonomous problems governed by evolution/differential inclusions in infinite dimensions. It is based on the construction of discrete approximations that largely exploits the *Lipschitzian* nature of the finitely many equality and inequality endpoint constraints imposed in (P_0) and then on passing to the limit from discrete approximations with taking into account specific features of subgradients of Lipschitzian functions. Implementing this scheme in the case of the delay-differential systems under consideration and employing the above developments of this paper for the delayed inclusions, we arrive at the necessary optimality conditions for (P_0) formulated in the theorem.

The *second scheme* exploited in the proof of Corollary 6.24 in Mordukhovich (2006b) for the case of nondelayed systems is based on employing *SNC calculus* results for endpoint constraint sets given by *finitely many* equalities and inequalities. It is proved in fact that such sets *do exhibit* the SNC property (which is strongly related to the more conventional *finite codimension property* in this setting) under some *qualification conditions* that are extensions of the classical Mangasarian-Fromovitz constraint qualification to the case of Lipschitzian functions. On the other hand, the *absence* of the afore-mentioned constraint qualification (when the SNC property may be violated) leads us to the *abnormal case* of the transversality inclusion also covered by the theorem. ■

References

BORWEIN, J.M. and ZHU, Q.J. (2005) *Techniques of Variational Analysis.* CMS Books in Mathematics, **20**, Springer, New York.

DIESTEL, J. and UHL, J.J. (1977) *Vector Measures.* American Mathematical Society, Providence, R.I.

DONTCHEV, A.L. and FARKHI, E.M. (1989) Error estimates for discreti-zed differential inclusions. *Computing*, **41**, 349–358.

MORDUKHOVICH, B.S. (1995) Discrete approximations and refined Euler-Lagrange conditions for nonconvex differential inclusions. *SIAM Journal on Control and Optimization*, **33**, 882–815.

MORDUKHOVICH, B.S. (2006a) *Variational Analysis and Generalized Differentiation, I: Basic Theory.* Grundlehren Series (Fundamental Principles of Mathematics Sciences), **330**, Springer, Berlin.

MORDUKHOVICH, B.S. (2006b) *Variational Analysis and Generalized Differentiation, II: Applications.* Grundlehren Series (Fundamental Principles of Mathematics Sciences), **331**, Springer, Berlin.

MORDUKHOVICH, B.S. (2007) Variational analysis of evolution inclusions. *SIAM Journal on Optimization*, **18**, 752–777.

MORDUKHOVICH, B.S. and WANG, D. (2005) Optimal control of semilinear evolution inclusions via discrete approximations. *Control and Cybernetics*, **34**, 849–870.

MORDUKHOVICH, B.S. and WANG, L. (2003) Optimal control of constrained delay-differential inclusions with multivalued initial conditions, *Control and Cybernetics*, **28**, 585–609.

MORDUKHOVICH, B.S. and WANG, L. (2004) Optimal control of neutral functional-differential inclusions, *SIAM Journal on Control and Optimization*, **43**, 111–136.

SCHIROTZEK, W. (2007) *Nonsmooth Analysis.* Series Universitext, Springer, Berlin.

SMIRNOV, G.V. (2002) *Introduction to the Theory of Differential Inclusions.* American Mathematical Society, Providence, R.I.

Tolstonogov, A.A. (2000) *Differential Inclusions in a Banach Space.* Kluwer, Dordrecht, The Netherlands.

Vinter, R.B. (2000) *Optimal Control.* Birkhäuser, Boston.

Warga, J. (1972) *Optimal Control of Differential and Functional Equations.* Academic Press, New York.