



On Generalized Bolza Problem and Its Application to Dynamic Optimization

Alexander D. Ioffe¹

Received: 19 October 2018 / Accepted: 1 February 2019 / Published online: 20 February 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

We consider two classes of problems: unconstrained variational problems of Bolza type and optimal control problems with state constraints for systems governed by differential inclusions, both under fairly general assumptions, and prove necessary optimality conditions for both of them. The proofs using techniques of variational analysis are rather short, compared to the existing proofs, and the results seem to cover and extend the now available. The key step in the proof of the necessary conditions for the second problem is an equivalent reduction to one or a sequence of reasonably simple versions of the first.

Keywords Optimal control · Differential inclusion · Maximum principle · Variational analysis · Subdifferential calculus

Mathematics Subject Classification 49K21 · 49J53 · 49J52 · 58C06

1 Introduction

The ultimate purpose of the paper is to study optimal control problems with state constraints for systems governed by differential inclusions. This is a fairly nontrivial class of optimization problems. The result we prove seems to offer the sharpest and most general necessary optimality conditions. It strengthens Clarke's recent theorem [1] (so far the strongest for problems without state constraints), in particular, by extending it to problems involving state constraints, and accordingly weakens the assumptions on which the principal results presented in the monograph by Vinter [2] (for problems with state constraints) are based.

But I believe the way the result has been obtained deserves a special, if not the main, attention. The key element of the proof is reduction of the original problem to uncon-

✉ Alexander D. Ioffe
ioffe@math.technion.ac.il

¹ Department of Mathematics, Technion, 32000 Haifa, Israel

strained minimization of a functional that looks similar to the classical Bolza problem of calculus of variations with Lipschitz integrand and off-integral term. (Actually, in the absence of state constraints it is a standard Bolza problem with nonsmooth integrand and off-integral term. In the presence of state constraints, the latter has more complicated structure.) The proof of necessary optimality conditions for strong minimum in this “generalized Bolza problem” is a combination of a simple relaxation theorem that allows to reduce the task to finding necessary conditions for weak minimum in the relaxed problem with subsequent application of standard rules of the nonconvex subdifferential calculus which is a well-developed chapter of local variational analysis.

The idea behind the reduction is actually very simple: if in an optimization problem the cost function is Lipschitz near a solution and constraints are subregular at the point, then the problem admits exact penalization with a Lipschitz (but necessarily nonsmooth) penalty function giving a certain measure of violation of constraints. This fact was first mentioned and used in [3], but the very idea to use one or another type of exact penalization appeared even earlier, at the very dawn of variational analysis. Ioffe and Tikhomirov in [4] and Rockafellar in [5] used such reduction to prove the existence of solutions in optimal control problems. Clarke in [6] applied similar reduction to prove the maximum principle for optimal control problems under the calmness assumption. This is a fairly weak assumption (implied, in particular, by metric regularity of constraints), although difficult to verify directly. However, the techniques used in [6] did not work in the absence of calmness and a different approach (associated with controllability) was chosen in the accompanying paper [7] to prove a full version of the maximum principle.

In the three mentioned works, the reduction was based on the standard trick when a mapping was replaced by the indicator function of its graph, so the penalty functions appeared to be extended real-valued. Loewen [8] was the first to use a Lipschitz penalty to reduce an optimal control problem with free end points (automatically calm) to a Bolza problem with everywhere finite integrand, even having certain Lipschitz properties. His proof of the maximum principle for such problems was amazingly simple. But again, the techniques could not be applied to problems with fixed end points and, like Clarke, Loewen had to use another technique for the proof of a general maximum principle.

A proof fully based on reduction to an unconstrained Bolza problem was given by Ioffe in [9] with the help of different exact penalization techniques associated with subregularity property of constraints. This technique was later formalized in [17] in terms of a certain “optimality alternative” (see the last section). The other key factor that made the proof in [9] possible was the necessary condition for Bolza problems obtained by Ioffe and Rockafellar in [10]. This paper was a crucial step in another (but associated) line of developments related to the very structure of necessary optimality conditions in optimal control problems with differential inclusions.

Here is a brief account of the developments up to the mentioned Clarke’s work [1]. Starting with his early papers [6,7] and up to Loewen’s book [8], the adjoint Euler–Lagrange inclusions were stated in terms of Clarke’s normal cones which are convex closures of limiting normal cones. In the pioneering papers by Smirnov [11] and Loewen and Rockafellar [12], a new and substantially more precise version of the Euler–Lagrange inclusion was introduced that involved only partial convexification

of the limiting normal cones. The problem was that in both cases the proof needed convexity of values of the set-valued mapping in the right-hand side of the inclusion. Mordukhovich in [13] showed that a modification of Smirnov's proof (for bounded-valued and Hausdorff Lipschitz set-valued maps) allows to establish the same type of Euler–Lagrange inclusion without the convexity assumption. However, the maximum principle was lost in this proof. In 1997, Ioffe in the mentioned paper [9] and Vinter and Zheng [14], using a different approach, did prove that both the partially convexified Euler–Lagrange inclusion and the maximum principle are necessary for a strong local minimum in optimal control problems with differential inclusions under fairly general conditions, basically the same as in [12] but without the convexity assumption on the set-valued map. A detailed proof of the result combining the approaches of [9, 14] is present in Vinter's monograph [2].

Another line of developments concerns Lipschitz properties of the set-valued mapping in the right-hand side of the differential inclusion near the optimal trajectory (see the comments after the statement of Theorem 3.3).

The plan of the paper is the following. The next section contains information, mainly from variational analysis, which is necessary for the statements and proofs of the main results. Section 3 is central. It contains the statements of the problems and the main theorems. We also give in this section some applications and versions of the theorems. The subsequent two sections contain the proofs of the main theorems for the generalized Bolza problem (Sect. 4) and the optimal control problem (Sect. 5), and in a short last section we return to a question that has remained open for quite a while. To conclude, we mention that for optimal control problems in the classical Pontryagin form the proofs can be noticeably simplified. We hope to discuss this in detail elsewhere.

2 Preliminaries

In what follows $x \in \mathbb{R}^n$, $\|x\|$ is the standard Euclidean norm of x , B is the unit ball, $B(x, r)$ is the closed ball of radius r around x and $\langle w, x \rangle$ stands for the inner product. The symbol $C[0, T]$ stands for the space of \mathbb{R}^n -valued continuous functions on $[0, T]$ with the standard norm

$$\|x(\cdot)\| = \max_{0 \leq t \leq T} \|x(t)\|;$$

by $W^{1,p}[0, T]$, $p \in [1, \infty)$ (or just $W^{1,p}$), we denote the space of \mathbb{R}^n -valued absolutely continuous functions on $[0, T]$ whose derivatives belong to L^p with the norm

$$\|x(\cdot)\|_{1,p} = \|x(0)\| + \|\dot{x}(\cdot)\|_p.$$

By $f \circ G$, we denote the composition of a mapping G and a function. But when dependence on the variables should be emphasized, we would often write $f(G(x))$ rather than $(f \circ G)(x)$.

2.1 Subdifferentials

We shall need several types of subdifferentials: proximal and limiting subdifferential in \mathbb{R}^n [2,15] and in general Hilbert spaces [16], Dini-Hadamard subdifferential in separable Banach spaces, and G -subdifferential [17,18] and Clarke’s generalized gradients in Banach spaces [19], all coinciding with the subdifferential in the sense of convex analysis for convex functions. We shall denote by ∂_p the proximal subdifferential, by ∂_C the generalized gradient of Clarke and by ∂ both the limiting subdifferential in \mathbb{R}^n and the subdifferential in the sense of convex analysis, by ∂^- the Dini-Hadamard subdifferential and by ∂_G the G -subdifferential in the general Banach space.

Here are some basic facts about these subdifferentials we shall need in proofs.

Proposition 2.1 ([10], Theorem 2) *Let X be a Hilbert space, and let f_1, \dots, f_k be lsc functions on X which are finite at a certain x and uniformly lower semicontinuous at x in the following sense: there is a $\varepsilon > 0$ such that for any sequences $\{x_{im}\}$ of elements of $B(x, \varepsilon) \cap \text{dom } f_i$ such that $\|x_{im} - x_{jm}\| \rightarrow 0$ as $m \rightarrow \infty$ for $i, j = 1, \dots, k$, there is a sequence $\{u_m\}$ such that $\|x_{im} - u_m\| \rightarrow 0$ for any $i = 1, \dots, k$ and*

$$\liminf_{m \rightarrow \infty} \sum_i (f_i(x_{im}) - f_i(u_m)) \geq 0.$$

Set $f(x) = f_1(x) + \dots + f_k(x)$ and let $x^ \in \partial_p f(x)$. Then for any $\delta > 0$ there are x_i with $\|x_i - x\| \leq \delta$ and $|f_i(x_i) - f_i(x)| < \delta$, and $x_i^* \in \partial_p f_i(x_i)$ such that $\|x_1^* + \dots + x_k^* - x^*\| < \delta, i = 1, \dots, k$.*

The uniform lower semicontinuity property is satisfied, in particular, when all functions but possibly one of them are Lipschitz near x .

Proposition 2.2 ([16], Chapter 3, Theorem 5.10) *Let $\varphi(t, x)$ be a function on $[0, T] \times \mathbb{R}^n$ which is measurable in t and lower semicontinuous in x . Consider in $L^2(0, T)$ the functional*

$$f(x(\cdot)) = \int_0^T \varphi(t, x(t))dt$$

and assume that $\xi(\cdot) \in \partial_p f(x(\cdot))$. Then $\xi(t) \in \partial_p \varphi(t, \cdot)(x(t))$ almost everywhere.

Note that it is assumed in [16] that φ does not depend on t and is globally Lipschitz as a function of x . But neither of the assumptions actually plays any role in the proof.

Proposition 2.3 ([17], Corollary 7.15) *Let X be a Banach space, and let the functions $f_i, i = 1, \dots, k$ be Lipschitz continuous in a neighborhood of a certain \bar{x} . Set $f(x) = \max_i f_i(x)$ and let $I = \{i : f_i(\bar{x}) = f(\bar{x})\}$. Then $\partial_G(f(\bar{x})) \subset \sum \alpha_i \partial_G f_i(\bar{x})$ over all $\alpha_i \geq 0$ such that $\alpha_i = 0$ for $i \notin I$ and $\sum \alpha_i = 1$.*

Proposition 2.4 ([17], Proposition 4.60) *Let X and Y be Banach spaces and $A : X \rightarrow Y$ a linear bounded operator with $\text{Im } A = Y$. Let further f be a function on Y which is Lipschitz continuous in a neighborhood of a certain $\bar{y} = A\bar{x}$. Set $g = f \circ A$. Then*

$$\partial_G g(\bar{x}) = A^*(\partial_G f(\bar{y})).$$

Surprisingly the next result seems to be absent in the literature, so we supply it with a complete proof (in which we use without explanation the definition of the G -subdifferential in a separable Banach space).

Proposition 2.5 *Let X and Y be separable Banach spaces, let $A : X \rightarrow Y$ be a bounded linear operator, and let φ be a function on Y which is Lipschitz continuous in a neighborhood of a certain $\bar{y} = A\bar{x}$. Set $f(x) = \varphi(Ax)$. Then*

$$\partial_G f(\bar{x}) \subset A^* \partial_G \varphi(\bar{y}).$$

Proof Take an x sufficiently close to \bar{x} , and suppose that $x^* \in \partial^- f(x)$. Setting $y = Ax$, we deduce that for all $h \in X$ and $v \in Y$

$$\langle x^*, h \rangle \leq \varphi^-(y; Ah) \leq \varphi^-(y, v) + K \|Ah - v\|,$$

where K is the Lipschitz constant of φ near \bar{y} , and therefore, the global Lipschitz constant of the lower Dini derivative of φ at y :

$$\varphi^-(y; v) = \liminf_{t \rightarrow 0} t^{-1}(\varphi(y + tv) - \varphi(y)).$$

Set $g(u, z) = K \|Au - z\|$. As $y = Ax$, we see that $g^-(x, y; (h, v)) = K \|Ah - v\|$ for all h, v . Note also that g is a convex continuous function. Hence, its lower directional derivative coincides with the usual directional derivative and the inequality above can be rewritten as

$$(x^*, 0) \in \partial^-(\varphi + g)(x, y).$$

We also have the obvious equality for the subdifferential of g :

$$\partial g(u, y) = \{(A^* v^*, -v^*) : v^* \in K \partial \|\cdot\|(Au - y)\}.$$

Both φ and g are Lipschitz functions, so we can apply the fuzzy sum rule (see, e.g., [17], Proposition 4.33) and find, given $\varepsilon > 0$ and a weak*-neighborhoods U and V of zeros in X and Y , some $u \in X, z_1, z_2 \in Y$ ε -close, respectively, to x and y , and $y^* \in \partial^- \varphi(z_1), (u^*, y_2^*) \in \partial g(u, z_2)$ such that the inclusions $u^* = A^* v^* \in x^* + U$ and $y^* - v^* \in V$ hold for some v^* with $\|v^*\| \leq K$. As both φ and g are Lipschitz, we can be sure that actually $u^* \in x^* + U \cap (\rho B)$ and $y^* - v^* \in V \cap (\rho B)$ for a sufficiently large ρ that does not depend on the choice of U and V . In other words, $A^* v^* = x^* + w^*$ and $y^* - v^* = z^*$ for some $w^* \in U$ and $z^* \in V$ whose norms do not exceed ρ .

Assume now that $x^* \in \partial_G f(\bar{x})$. Then there are x_m converging to \bar{x} and $x_m^* \in \partial f^-(x_m)$ weak* converging to x^* . Let now x_1, x_2, \dots and y_1, y_2, \dots be dense countable subsets of X and Y . Take $U_m = \{u^* \in X^* : |\langle u^*, x_i \rangle| < m^{-1}, i = 1, \dots, m\}$ and $V_m = \{z^* : |\langle z^*, y_i \rangle| < m^{-1}, i = 1, \dots, m\}$. Let finally u_m, z_{1m}, z_{2m} and

$u_m^*, y_m^*, v_m^*, w_m^*, z_m^*$ satisfy the above relations with $x = x_m, U = U_m$ and $V = V_m$, that is, $y_m = Ax_m$ and

$$\begin{aligned} &\|u_m - x_m\| < m^{-1}, \quad \|z_{1m} - y_m\| < m^{-1}, \quad \|z_{2m} - y_m\| < m^{-1}; \\ &u_m^* = A^*v_m^*, \quad w_m^* = u_m^* - x_m^* \in U_m; \\ &y_m^* \in \partial^-\varphi(z_{1m}), \quad (u_m^*, z_m^*) \in \partial g(u_m, z_{2m}), \quad y_m^* - v_m^* = z_m^* \in V_m. \end{aligned}$$

Then $u_m \rightarrow \bar{x}, z_{im} \rightarrow \bar{y}$ and all sequences of the functionals, being bounded, can be assumed weak* converging. Obviously, w_m^* and z_m^* converge to zero by the Banach–Steinhaus theorem. Therefore, u_m^* weak* converges to x^* ; y_m^* and v_m^* weak* converge to some y^* . It follows that $y^* \in \partial_G\varphi(\bar{y})$ and $x^* = A^*v^*$ as claimed.

Remark 2.1 If A is an operator onto Y , the proposition remains valid if φ is just lower semicontinuous (and finite at \bar{x}). Indeed, in this case, given $x \in X$ and $v \in Y$, there is a $u \in X$ such that $Au = v$ and $\|x - u\| \leq K\|A(x - u)\|$, where K is any number greater than the inverse of the Banach constant of A (the maximal radius of a ball in Y covered by the A -image of the unit ball in X). It follows that

$$d((x, \alpha), \text{epi } f) \leq Kd((Ax, \alpha), \text{epi } \varphi) \leq K((d(y, \alpha), \text{epi } \varphi) + \|y - Ax\|)$$

for any $(x, y, \alpha) \in X \times Y \times \mathbb{R}$, and the result follows from Theorem 7.19 of [17].

Remark 2.2 The previous results remain valid for a function $g = f \circ F$ with a continuously differentiable $F : X \rightarrow Y$ if as A we take the derivative of F at the point of interest.

2.2 Lipschitz Properties of Set-Valued Mappings

We denote by $F : Q \rightrightarrows P$ a set-valued mapping from Q to P . If both Q and P are topological spaces, then F is said to be a *closed mapping* if its graph is closed in the product topology. The expressions like “closed-valued mapping,” “convex-valued mapping,” etc., need no explanation.

Speaking about Lipschitz properties of maps, we of course assume that both the domain and range spaces are at least metric. Here we mainly deal with mappings between Euclidean spaces and shall use the corresponding notation and language starting with this point. The simplest and the most convenient Lipschitz property of a set-valued map to work with occurs when the mapping is Lipschitz with respect to the Hausdorff metric in the range space. This, however, is a very restrictive property, especially when the values of the mapping are unbounded sets.

A weaker property, that also has an advantage of being local, was introduced by Aubin: it is said that F is *pseudo-Lipschitz*, or has the *Aubin property* at $(\bar{x}, \bar{y}) \in \text{Graph } F$ if there are $r > 0, R > 0$ and $\varepsilon > 0$ such that

$$F(x) \cap B(\bar{y}, r) \subset F(x') + R\|x - x'\|B, \quad \forall x, x' \in B(\bar{x}, \varepsilon).$$

We shall call R a *Lipschitz constant* of F at (\bar{x}, \bar{y}) . The following simple proposition offers a nice metric characterization of the Aubin property.

Proposition 2.6 *Let $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ and $(\bar{x}, \bar{y}) \in \text{Graph } F$. Then the following two statements are equivalent:*

- (i) *F has the Aubin property at (\bar{x}, \bar{y}) ;*
- (ii) *there are $r > 0$, $R > 0$ and $\varepsilon > 0$ such that the function $x \rightarrow d(y, F(x))$ is R -Lipschitz in the ε -ball around \bar{x} for any y with $\|y - \bar{y}\| \leq r$.*

The implication (ii) \rightarrow (i) is straightforward. The opposite implication is well known (see, e.g., [15], Exercise 9.37). Note also that the ε and R in both properties coincide, while the r in (ii) may be smaller.

We shall also use a noninfinitesimal version of the pseudo-Lipschitz property with a priori fixed constants r , R and ε . The following fundamental result shows that a part of the mapping associated with the property admits lifting to a bounded-valued mapping which is Lipschitz with respect to the Hausdorff metric.

Proposition 2.7 *Let $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ be a set-valued mapping with closed graph. Let an \bar{x} and a closed set $C \subset F(\bar{x})$ be given such that for some $r > 0$, $R > 0$, $\varepsilon > 0$ and $\eta > 0$ the relations*

$$F(x) \cap (u + rB) \subset F(x') + R\|x - x'\|B, \quad F(x) \cap B(u, (1 - \eta)r) \neq \emptyset$$

hold for any $u \in C$ and $x, x' \in B(\bar{x}, \varepsilon)$. Set

$$\Gamma(x) = \{(\lambda, y) : \lambda \in [0, 1], y = u + \lambda(z - u), u \in C, z \in F(x), \|z - u\| \leq (1 - \lambda\eta)r\}.$$

Then there is a $c = c(\eta, r)$ such that Γ is cR -Lipschitz with respect to the Hausdorff metric in the ε -neighborhood of \bar{x} .

This is a slight generalization of a Clarke’s result ([1], Lemma 1) in which C is a one point set. We omit the proof as it repeats almost word for word the original proof by Clarke.

Lipschitz continuity of $d(y, F(\cdot))$, on the other hand, offers a convenient way to compute normal cones to the graph of F which naturally appear in necessary optimality conditions in problems involving set-valued mappings.

Proposition 2.8 ([9], Proposition 1) *Let $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ be a closed-valued mapping having the Aubin property at some $(\bar{x}, \bar{y}) \in \text{Graph } F$. Then the (limiting) normal cone to $\text{Graph } F$ at (\bar{x}, \bar{y}) is generated by the limiting subdifferential of the function $(x, y) \rightarrow d(y, F(x))$ at (\bar{x}, \bar{y}) .*

The last result we have to mention offers a geometric characterization of the Aubin property.

Proposition 2.9 *Let $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ be a set-valued mapping with closed graph that has the Aubin property at $(\bar{x}, \bar{y}) \in \text{Graph } F$ with Lipschitz constant R . Let $(q, p) \in N(\text{Graph } F, (\bar{x}, \bar{y}))$. Then $\|q\| \leq R\|p\|$.*

This is a simple and well-known fact whose proof, however, does not seem to be explicitly available in the existing literature. The inequality is straightforward for Fréchet normal cones (or regular normals in terminology of [15]). The Fréchet normal cone to the graph of F at (x, y) is defined as the collection of pairs (q, p) such that $\langle q, h \rangle - \langle p, v \rangle \leq o(\|h\| + \|v\|)$ if h and v are sufficiently small and $y + v \in F(x + h)$. By the Aubin property, given an h , there is a v satisfying $\|v\| \leq R\|h\|$. If now we take $h = \lambda q$, then we get $\lambda\|q\|^2 \leq \langle p, v \rangle \leq \lambda R\|p\|\|q\| + o(\|\lambda q\|)$. It remains to take into account that $N(\text{Graph } F, (\bar{x}, \bar{y}))$ is the outer limit of Fréchet normal cones at $(x, y) \in \text{Graph } F$, when $(x, y) \rightarrow (\bar{x}, \bar{y})$.

2.3 Measurability

A set-valued mapping $F : [0, T] \rightrightarrows \mathbb{R}^m$ is *measurable* if its graph belongs to the sigma-algebra generated by all products $\Delta \times Q$, where $\Delta \subset [0, T]$ is Lebesgue measurable and $Q \subset \mathbb{R}^n$ is open. For closed-valued and open-valued mappings, a convenient equivalent definition can be given in terms of so-called *Castaing representation*: F is measurable if and only if there is a countable collection $(u_i(\cdot))$ of measurable selections of F such that for almost every t the set $\{u_i(t), i = 1, 2, \dots\}$ is dense in $F(t)$.

A set-valued mapping $F : [0, T] \times \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is called *measurable (in the standard sense)* if the set-valued mapping $t \rightarrow \text{Graph } F(t, \cdot)$ is measurable. An extended real-valued function f on $[0, T] \times \mathbb{R}^n$ is measurable (in the standard sense) if so is the mapping $\text{epi } f(t) = \{t, \alpha : \alpha \geq f(t)\}$.

The most important fact is that all basic operations used in analysis preserve measurability. For details, we refer to [15,17].

2.4 Relaxation

Consider the functional $\int_0^T L(t, x(t), \dot{x}(t))dt$, assuming that the integrand L is continuous in x and $L(t, x(t), u(t))$ is measurable whenever $x(\cdot)$ is continuous and $u(\cdot)$ is measurable.

Proposition 2.10 (relaxation theorem) *Given an $x(\cdot) \in W^{1,1}$ and summable \mathbb{R}^n -valued functions $u_1(\cdot), \dots, u_k(\cdot)$. Set $u_0(t) = \dot{x}(t)$ and assume that there are $\varepsilon > 0$ and a summable $c(t)$ such that for all $i = 0, \dots, k$*

$$\max_{\|x - x(t)\| \leq \varepsilon} |L(t, x, u_i(t))| \leq c(t), \quad \text{a.e..}$$

Then there is a $\delta > 0$ such that for any $\alpha_i \geq 0, i = 1, \dots, k, \sum \alpha_i < \delta$ there is a sequence of $x_m(\cdot) \in W^{1,1}$ such that $\dot{x}_m(t) \in \{\dot{x}(t), u_1(t), \dots, u_k(t)\}$ for almost every $t, x_m(\cdot)$ weakly in $W^{1,1}$ converge to

$$w(t) = x(t) + \sum_i \alpha_i \int_0^t (u_i(\tau) - \dot{x}(\tau))d\tau \tag{1}$$

and

$$\begin{aligned} \limsup_{m \rightarrow \infty} \int_0^T L(t, x_m(t), \dot{x}_m(t)) dt &\leq \int_0^T L(t, w(t), \dot{x}(t)) dt \\ &+ \sum_i \alpha_i \int_0^T (L(t, w(t), u_i(t)) - L(t, w(t), \dot{x}(t))) dt. \end{aligned} \tag{2}$$

This is a simplified version of Bogolyubov’s convexification theorem [20] (see also [4] for more details).

Proof Take $\delta < 1$ to guarantee that $\delta \|u_i(\cdot) - \dot{x}(\cdot)\|_{L^1} < \varepsilon/2$ for all $i = 1, \dots, k$. Let α_i and $x(\cdot)$ satisfy the conditions with the chosen δ .

Now given an m , set $\Delta_j = [(j - 1)/m, j/m]$, $j = 1, \dots, m$, and let Δ_{ji} , $i = 1, \dots, k$ be subsets of Δ_j with Lebesgue measure α_i/m , respectively, such that $\Delta_{ji} \cap \Delta_{j'i'} = \emptyset$ if $i \neq i'$. Define $x_m(\cdot)$ by

$$x_m(0) = x(0); \quad \dot{x}_m(t) = \begin{cases} u_i(t), & \text{if } t \in \Delta_{ji} \text{ for some } j, i; \\ \dot{x}(t), & \text{otherwise.} \end{cases}$$

It is clear that $\dot{x}_m(\cdot)$ are uniformly integrable, so we may assume that $x_m(\cdot)$ converge uniformly to some $w(\cdot)$. We can rewrite the definition of \dot{x}_m as follows:

$$\dot{x}_m(t) = \left(1 - \sum_{i=1}^k \alpha_{im}(t) \right) \dot{x}(t) + \sum \alpha_{im}(t) u_i(t),$$

where

$$\alpha_{im}(t) = \begin{cases} 1, & \text{if } t \in \Delta_{ji}, \text{ for some } j = 1, \dots, m; \\ 0, & \text{otherwise.} \end{cases}$$

Then

$$\begin{aligned} L(t, x_m(t), \dot{x}_m(t)) &= L(t, x_m(t), \dot{x}(t)) + \sum_{i=1}^k \alpha_{im}(t) (L(t, x_m(t), u_i(t)) \\ &\quad - L(t, x_m(t), \dot{x}(t))). \end{aligned}$$

It is clear that every $\alpha_{im}(\cdot)$ weak*-converges in L^∞ (and hence weakly in L^1) to the function identically equal to α_i . The result is now immediate. \square

3 Statements of the Problems and Main Results

The principal object to be studied in the paper is the *generalized Bolza problem*

$$\text{minimize } J(x(\cdot)) = \varphi(x(\cdot)) + \int_0^T L(t, x(t), \dot{x}(t))dt, \tag{P}$$

with φ being a function on the space $C[0, T]$, not necessarily a traditional function of the end points $(x(0), x(T))$.

In what follows, we shall fix some $\bar{x}(\cdot) \in W^{1,1}$ which will be assumed a local minimizer of J in one or another sense. Here are the basic hypotheses on the functions φ and L :

(H₁) φ is Lipschitz in a neighborhood of $\bar{x}(\cdot)$ in $C[0, T]$, that is, there are $K > 0$ and $\bar{\varepsilon} > 0$ such that

$$|\varphi(x(\cdot)) - \varphi(x'(\cdot))| \leq K \|x(\cdot) - x'(\cdot)\|;$$

if $x(\cdot), x'(\cdot)$ are $\bar{\varepsilon}$ -close to $\bar{x}(\cdot)$ in $C[0, T]$.

(H₂) $L(t, x, y)$ is lower semicontinuous in (x, y) and measurable in the standard sense¹ and there are $\bar{\varepsilon} > 0$, measurable $r(t) > 0$, summable $R(t) \geq 0$ and $k(t) \geq 0$ and a measurable set-valued mapping $D(\cdot) : [0, T] \rightrightarrows \mathbb{R}^n$ such that for almost every $t \in [0, T]$ the function $L(t, \cdot)$ is continuous on $B(\bar{x}(t), \bar{\varepsilon}) \times D(t)$ and

$$\begin{aligned} |L(t, x, y) - L(t, x', y)| &\leq R(t)\|x - x'\|, \quad \forall x, x' \in B(\bar{x}(t), \bar{\varepsilon}), y \in B(\bar{x}(t), r(t)); \\ |L(t, x, y)| &\leq k(t), \quad \forall x \in B(\bar{x}(t), \bar{\varepsilon}), y \in B(\bar{x}(t), \bar{\varepsilon}). \end{aligned} \tag{3}$$

hold a.e. on $[0, T]$.

Note that the conditions do not exclude the possibility that L is extended real-valued, in particular, that $L(t, x, y) = \infty$ if $y \notin D(t)$, even if x is close to $\bar{x}(t)$. It should also be taken into account that $\inf r(t)$ can be zero and L need not be Lipschitz with respect to the last argument.

Theorem 3.1 *Assume (H₁) and (H₂). If $\bar{x}(\cdot)$ is a local minimizer of J on $W^{1,1}$, then there are an \mathbb{R}^n -valued Radon measure $\nu \in \partial_G \varphi(\bar{x}(\cdot))$, an \mathbb{R}^n -valued function of bounded variations $p(t)$, continuous from the left and a summable \mathbb{R}^n -valued function $q(t)$ such that*

$$p(t) = - \int_t^T q(s)ds - \int_t^T d\nu(t), \quad p(0) = 0, \tag{4}$$

and the following conditions are satisfied almost everywhere on $[0, T]$:

$$q(t) \in \text{conv} \{q : (q, p(t)) \in \partial L(t, \cdot, \cdot)(\bar{x}(t), \dot{\bar{x}}(t))\} \tag{5}$$

¹ Actually, it will be sufficient to assume that $L(t, x(t), y(t))$ is measurable if so are $x(t)$ and $y(t)$

(the Euler inclusion) and

$$L(t, \bar{x}(t), u) - L(t, \bar{x}(t), \dot{\bar{x}}(t)) - \langle p(t), u - \dot{\bar{x}}(t) \rangle \geq 0 \quad \text{if } u \in D(t) \tag{6}$$

(the Weierstrass condition).

An interesting (and perhaps the most interesting) case corresponds to $D(t) \equiv \mathbb{R}^n$. It will be easy to see from the proof that in this case the Euler inclusion (5) is a necessary condition for a weak minimum in (\mathbf{P}) , that is, on a set of $x(\cdot)$ that are $W^{1,\infty}$ -close to $\bar{x}(\cdot)$. Looking ahead, we just say that the proof of this fact is noticeably simpler and need not any reference to the relaxation theorem.

We deduce from the theorem in this case that the Euler inclusion (5) and the Weierstrass condition (6) together give a necessary condition for a strong minimum in the classical sense, that is, when there is an $\varepsilon > 0$ such that $J(x(\cdot)) \geq J(\bar{x}(\cdot))$ for all $x(\cdot) \in W^{1,1}$ satisfying $\|x(t) - \bar{x}(t)\| < \varepsilon, \forall t$. Another point we wish to mention in this connection is that $L(t, x, y)$ have to satisfy the Lipschitz property w.r.t x only for y in a possibly small neighborhood of $\dot{\bar{x}}(t)$.

Applying the theorem to the nonsmooth version of the standard Bolza problem:

$$\text{minimize } \ell(x(0), x(T)) + \int_0^T L(t, x(t), y(t))dt \tag{P_{st}}$$

(under the assumption that $D(t) \equiv \mathbb{R}^n$), we get the following extension of the classical necessary conditions.

Theorem 3.2 Assume that ℓ is Lipschitz near $(\bar{x}(0), \bar{x}(T))$ and L satisfies (H_2) with $D(t) \equiv \mathbb{R}^n$. If $\bar{x}(\cdot)$ is a weak local minimum in the problem, then there is a $p(\cdot) \in W^{1,1}$ satisfying the Euler inclusion

$$\dot{p}(t) \in \text{conv} \{q : (q, p(t)) \in \partial L(t, \cdot, \cdot)(\bar{x}(t), \dot{\bar{x}}(t))\} \tag{7}$$

a.e. on $[0, T]$ and the end points condition

$$(p(0), -p(T)) \in \partial \ell((\bar{x}(0), \bar{x}(T))). \tag{8}$$

If, moreover, $\bar{x}(\cdot)$ is a strong minimum, then for almost every t the Weierstrass condition (4) holds for all $u \in \mathbb{R}^n$ with the same $p(\cdot)$.

Proof As $\varphi(x(\cdot)) = \ell(x(0), x(T))$ is a function of terminal values of $x(\cdot)$, it is an easy matter to see that $v \in \partial_G \varphi(x(\cdot))$ if and only if there are $g, h \in \mathbb{R}^n$ such that $(g, h) \in \partial \ell(x(0), x(T))$ and the action of v on elements of $C([0, T])$ is defined by

$$\int_0^T u(t)v(dt) = \langle h, u(T) \rangle + \langle g, u(0) \rangle.$$

Applying Theorem 3.1, we therefore find some $(g, h) \in \partial \ell((\bar{x}(0), \bar{x}(T)))$ such that for some $q(t)$ and $p(t) = - \int_t^T v(dt) - \int_t^T q(s)ds = -h - \int_t^T q(s)ds$ (for $t > 0$) we

get (if we redefine $p(t)$ at zero by continuity: $p(0) = -h + \int_0^T q(t)dt$) $p(0) - g = 0$, hence (8).

Let us now turn to the optimal control problem:

$$\begin{aligned} &\text{minimize } \ell(x(0), x(T)), \\ &\text{s.t. } \dot{x} \in F(t, x), \quad (x(0), x(T)) \in S, \quad g(t, x(t)) \leq 0, \quad \forall t. \end{aligned} \tag{OC}$$

Here we assume that there are $\bar{x}(\cdot) \in W^{1,1}$ and $\bar{\varepsilon} > 0$ such that

- (H₃) ℓ is Lipschitz near $(\bar{x}(0), \bar{x}(T))$, $S \subset \mathbb{R}^n \times \mathbb{R}^n$ is closed;
- (H₄) g is upper semicontinuous on the set $\{(t, x) : t \in [0, T], \|x - \bar{x}(t)\| \leq \bar{\varepsilon}\}$ and there is a $K > 0$ such that for $x, x' \in B(\bar{x}(t), \bar{\varepsilon})$

$$|g(t, x) - g(t, x')| \leq K\|x - x'\| \quad \text{a.e. on } [0, T].$$

- (H₅) F is closed-valued and measurable in the standard sense and there are a measurable $\bar{r}(t) > 0$ bounded away from zero, a summable $\bar{R}(t) \geq 0$ and an $\eta \in]0, 1[$ such that the relations

$$\begin{aligned} F(t, x) \cap B(\dot{\bar{x}}(t), \bar{r}(t)) &\subset F(t, x') + \bar{R}(t)\|x - x'\|B, \\ F(t, x) \cap B(\dot{\bar{x}}(t), (1 - \eta)\bar{r}(t)) &\neq \emptyset \end{aligned} \tag{9}$$

hold for all $x, x' \in B(\bar{x}(t), \bar{\varepsilon})$.

The first of the relations is the a nonlocal extension of Aubin’s pseudo-Lipschitz property. In one or another form, it was used in all results on the maximum principle for differential inclusions starting with [12]. The necessity of the second condition in (9) under weaker versions of the first was observed by Clarke in [1].

We need more notation to state the theorem. Let $U(t)$ be the closure of the collection of all $u \in F(t, \bar{x}(t))$ such that

$$\begin{aligned} F(t, x) \cap B(u, r) &\subset F(t, x') + R\|x - x'\|B, \quad F(t, x) \cap B(u, (1 - \eta)r) \neq \emptyset, \tag{9'} \\ &\text{if } \|x - \bar{x}(t)\| \leq \varepsilon \end{aligned}$$

for some $r > 0, R > 0, \eta > 0, \varepsilon > 0$ (depending on u). Finally, set $\Delta(x(\cdot)) = \{t : g(t, x(t)) = \max_t g(t, x(t))\}$ and

$$\partial_C^> g(t, x) = \text{conv} \{\limsup \partial g(t_m, x_m) : t_m \rightarrow t, x_m \rightarrow x, g(t_m, x_m) > g(t, x)\}.$$

Theorem 3.3 *We posit (H₃)–(H₅) and assume that $\bar{x}(t) \in W^{1,1}$ is a local solution in (OC). Then there are $\lambda \geq 0$, a nonnegative measure μ on $[0, T]$ supported on $\Delta(\bar{x}(\cdot))$, a measurable \mathbb{R}^n -valued function $q(t)$ satisfying $\|q(t)\| \leq R(t)$ a.e., an \mathbb{R}^n -valued function $p(t)$ of bounded variation, continuous from the left, and a μ -measurable selection $\gamma(\cdot)$ of the set-valued mapping $t \rightarrow \partial_C^>(t, x(t))$ such that $dp(t) = q(t)dt + \gamma(t)\mu(dt)$ and the following four conditions are satisfied:*

- (i) $\lambda + \|p(\cdot)\| + \mu([0, T]) = 1$ (nontriviality);
- (ii) $(p(0), -p(T) - \gamma(T)\mu(\{T\}) \in \lambda \partial \ell(\bar{x}(0), \bar{x}(T)) + N(S, (x(0), x(T)))$ (transversality);
- (iii) $q(t) \in \text{conv} \{q : (q, p(t)) \in N(\text{Graph } F(t, \cdot), (\bar{x}(t), \dot{\bar{x}}(t)))\}$ a.e. on $[0, T]$ (Euler–Lagrange inclusion);
- (iv) $\langle p(t), y - \dot{\bar{x}}(t) \rangle \leq 0, \forall y \in U(t)$ a.e. on $[0, T]$ (maximum principle).

By (H_5) , $F(t, \cdot)$ has the Aubin property at every $u \in F(t, \bar{x}(t))$ with $\|u - \dot{\bar{x}}(t)\| < \bar{r}(t)$. Thus, replacing in the statement $U(t)$ by $B(\bar{x}(t), \bar{r}(t))$, we get an extension of Clarke’s “stratified maximum principle” to optimal control problems with state constraints. It seems to be appropriate at this point to add a few words concerning the evolution of the assumptions concerning Lipschitz properties of F . In [7,8,11], the mapping was assumed globally Lipschitz in the state variable with the Lipschitz constant being a summable function of the time variable. In [12], this property was weakened and replaced by a certain global version of the Aubin pseudo-Lipschitz property with a linear growth of the Lipschitz constant as a function of the radius of the ball on which the mapping is considered. This assumption (also applied in [9,14]), although still sufficiently restrictive, made possible meaningful treatment of differential inclusions with unbounded right-hand sides. Finally, Clarke’s stratified maximum principle implied that arbitrary rate of growth works as well.

Remark 3.1 There are certain differences in formulations of the maximum principle for problems with state constraints in the literature. The above statement is structured along the lines of the maximum principle proved in [4] (Theorem 1 of § 5.2). On the other hand, the statement of the maximum principle for **(OC)** proved in [2] look differently at the first glance. But it is not a difficult matter to notice that the functions $p(t)$ here and $q(t)$ in [2] coincide up to a delicate difference: the first is continuous from the left and may be discontinuous at the left end of the interval while the second is continuous from the right and may be discontinuous at the right end of the interval.

4 Proof of Theorem 3.1

Without loss of generality, we can assume in the proof that $\bar{x}(t) \equiv 0$.

1. Let $u_i(\cdot), i = 1, \dots, k$ be bounded measurable selections of $D(t)$ a.e.. such that for some $\varepsilon > 0$ the functions $\max_{\|x - \bar{x}(t)\| \leq \varepsilon} L(t, x, u_i(t))$ is summable. Then $u_i(\cdot)$ satisfy the conditions of Proposition 2.10 with any $x(\cdot) \in W^{1,1}$ close to $\bar{x}(\cdot)$. It follows immediately from the proposition that the vector $(\bar{x}(\cdot), 0, \dots, 0)$ is a weak local minimizer of the functional

$$\hat{J}(x(\cdot), \alpha_1, \dots, \alpha_k) = \varphi(w(\cdot)) + \int_0^T L(t, w(t), \dot{x}(t))dt + \sum_i \alpha_i^+ \int_0^T (L(t, w(t), u_i(t)) - L(t, w(t), \dot{x}(t)))dt,$$

with $w(\cdot)$ defined by (1), subject to the constraint $\|\dot{x}(t) - \dot{\bar{x}}(t)\| \leq r(t)$ a.e..

(By saying that $(\bar{x}(\cdot), 0, \dots, 0)$ is a weak minimizer of \hat{J} , we mean that there is an $\varepsilon > 0$ such that $\hat{J}((x(\cdot), \alpha_1, \dots, \alpha_k)) \geq \hat{J}((\bar{x}(\cdot), 0, \dots, 0))$ whenever $\|x(\cdot)\|_{1,\infty} < \varepsilon$ and $0 \leq \alpha_i < \varepsilon$.)

Consider the space $Z = \mathbb{R}^n \times L^2 \times L^2 \times \mathbb{R}^k$ and let the operator $\Lambda : Z \rightarrow W^{1,2}$ associate with every $(a, x(\cdot), y(\cdot), \alpha) \in Z$, where $\alpha = (\alpha_1, \dots, \alpha_k)$, the function

$$\Lambda(a, y(\cdot), \alpha_1, \dots, \alpha_k)(t) = a + \int_0^t \left(y(\tau) + \sum_i \alpha_i (u_i(\tau) - y(\tau)) \right) d\tau.$$

To simplify notation, we can assume without loss of generality that $\bar{x}(t) \equiv 0$. Set

$$\tilde{L}(t, x, y) = \begin{cases} L(t, x, y), & \text{if } \|x\| \leq \bar{\varepsilon}, \|y\| \leq \min\{\bar{\varepsilon}, r(t)\}; \\ \infty, & \text{otherwise;} \end{cases}$$

$$g_{i\varepsilon}(t) = \max_{\|x\|, \|y\| \leq \varepsilon} (L(t, x, u_i(t)) - L(t, x, y));$$

and consider the following four functionals on Z (where of course $z = (a, x(\cdot), y(\cdot), \alpha_1, \dots, \alpha_k)$):

$$I_1(z) = \varphi(\Lambda(a, y(\cdot), \alpha_1, \dots, \alpha_k));$$

$$I_2(z) = \int_0^T \tilde{L}(t, x(t), y(t)) dt;$$

$$I_3(z) = \int_0^T R(t) \left\| x(t) - a - \int_0^t \left(y(\tau) + \sum_i \alpha_i (u_i(\tau) - y(\tau)) \right) d\tau \right\| dt;$$

$$I_4(z) = \sum_i \alpha_i \int_0^T g_{i\varepsilon}(t) dt + K \sum_i \alpha_i^-,$$

and set $I = I_1 + I_2 + I_3 + I_4$. It is an easy matter to see that $\hat{J}(0, \dots, 0) = I(0, \dots, 0)$ and (setting $w(t) = a + \int_0^t y(\tau) d\tau$)

$$\hat{J}(w(\cdot), \alpha_1, \dots, \alpha_k) \leq I(a, x(\cdot), y(\cdot), \alpha_1, \dots, \alpha_k).$$

if $\|a\| + \int_0^T \|y(t)\| dt \leq \bar{\varepsilon}$, $\|x(t)\| \leq \bar{\varepsilon}$ for all t , $\|y(t)\| \leq \min\{\bar{\varepsilon}, r(t)\}$ a.e., $\alpha_i \geq 0$ and K is sufficiently large. It follows that zero is a local minimum of $I(z)$ in Z . Therefore,

$$0 \in \partial_p I(0).$$

2. Analysis of the inclusion is the next step in the proof. The functionals I_j are lower semicontinuous and uniformly lower semicontinuous near zero in Z since

all I_j except I_2 satisfy the Lipschitz condition. By Proposition 2.1 for any given $\delta > 0$, there are $z_j = (a_j, x_j(\cdot), y_j(\cdot), \alpha_{1j}, \dots, \alpha_{kj}) \in Z$, and $z_j^* = (b_j, x_j^*, y_j^*, \beta_1, \dots, \beta_k) \in Z^*$, $j = 1, 2, 3, 4$ such that $|I_j(z_j) - I_j(0)| < \delta$ and

$$\begin{aligned}
 &(b_j, x_j^*, y_j^*, \beta_1, \dots, \beta_k) \in \partial_p I_j(z_j); \\
 &\|a_j\| + \|x_j(\cdot)\| + \|y_j(\cdot)\| + \sum_{i=1}^k |\alpha_{ij}| < \delta; \\
 &\left\| \sum_{j=1}^4 z_j^* \right\| = \max \left\{ \left\| \sum_{j=1}^4 b_j \right\|, \left\| \sum_{j=1}^4 x_j^* \right\|, \left\| \sum_{j=1}^4 y_j^* \right\|, \left| \sum_{j=1}^4 \beta_{1j} \right|, \dots, \left| \sum_{j=1}^4 \beta_{kj} \right| \right\} < \delta
 \end{aligned}
 \tag{10}$$

Finding estimates for $\partial_p I_j$ does not require much effort.

- Let $(b, x^*, y^*, \beta_1, \dots, \beta_k) \in \partial_G I_1(z)$. Then $x^* = 0$ as I_1 does not depend on $x(\cdot)$, and if $w(\cdot) = \Lambda(z)$, then there is a measure $\nu \in \partial_G \varphi(w(\cdot))$ such that

$$\begin{aligned}
 b &= \int_0^T \nu(dt), \quad \langle y^*, \nu(\cdot) \rangle = (1 - \sum \alpha_i) \int_0^T \langle v(t), v(t) \rangle dt, \\
 \beta_i &= \int_0^T \langle v(t), u_i(t) - y(t) \rangle dt,
 \end{aligned}$$

where we have set $\nu(t) = \int_t^T \nu(dt)$.

Indeed, I_1 can be viewed as a composition of Λ and the restriction of φ to $W^{1,2}$. Denote for a moment this restriction by $\tilde{\varphi}$. Λ is a smooth mapping and its derivative at zero (and hence at all nearby points) is onto. By Proposition 2.4, $\partial_G(z) = \Lambda'(z)^*(\partial_G \tilde{\varphi})(z)$. On the other hand, $\tilde{\varphi}$ is the composition of embedding $W^{1,2}$ into $C([0, T])$ and φ . Therefore, by Proposition 2.5 $\partial_G \tilde{\varphi}(x(\cdot))$ is contained in the set of restrictions of elements of $\partial_G \varphi(x(\cdot))$ to $W^{1,2} \times \mathbb{R}^k$. It remains to recall that the proximal subdifferential is contained in the G -subdifferential.

- If $(b, x^*, y^*, \beta_1, \dots, \beta_k) \in \partial_p I_2(z)$, then $b = 0$, all $\beta_i = 0$, and there are $\lambda(\cdot)$ and $\mu(\cdot)$ belonging to L^2 such that $(\lambda(t), \mu(t)) \in \partial_p \tilde{L}(t, \cdot)(x(t), y(t))$ almost everywhere and

$$\langle x_2^*, h(t) \rangle + \langle y_2^*, v(t) \rangle = \int_0^T (\langle \lambda(t), h(t) \rangle + \langle \mu(t), v(t) \rangle) dt$$

for all $h(\cdot)$ and $v(\cdot)$ in L^2 . This is immediate from Proposition 2.2.

- If $(b, x^*, y^*, \beta_1, \dots, \beta_k) \in \partial_p I_3(z)$, then there is a $\xi(\cdot) \in L^2$ with $\|\xi(t)\| \leq R(t)$ a.e. such that

$$\begin{aligned}
 b &= - \int_0^T \xi(t) dt, \quad \langle x^*, h(\cdot) \rangle = \int_0^T \langle \xi(t), h(t) \rangle dt, \\
 \langle y^*, v(\cdot) \rangle &= - \left(1 - \sum_i \alpha_i \right) \int_0^T \langle \eta(t), v(t) \rangle dt, \quad \beta_i = - \int_0^T \langle \eta(t), u_i(t) - y(t) \rangle dt,
 \end{aligned}$$

where $\eta(t) = - \int_t^T \xi(t) dt$.

Indeed, I_3 is a composition of a convex lsc functional $\int_0^T R(t) \|x(t)\| dt$ and a smooth mapping from Z into L^2 . For such mappings, all subdifferentials coincide and are equal to the composition of the derivative of the inner mapping and the convex subdifferential of the outer function.

– if $(b, x^*, y^*, \beta_1, \dots, \beta_k) \in \partial_p I_4(z)$, then there are $\rho_i \in [-K, 0]$ such that

$$b = 0, \quad x^* = 0, \quad y^* = 0, \quad \beta_i = \int_0^T g_{i\varepsilon}(t) dt + \rho_i.$$

Taking (10) into account, we conclude that for some $z_j, j = 1, 2, 3, 4$, satisfying the second inequality in (10) there are a regular \mathbb{R}^n -valued Radon measure $\nu \in \partial_G \varphi(z_1)$, measurable $(\lambda(t), \mu(t)) \in \partial_p \tilde{L}(t, \cdot, \cdot)(x_2(t), y_2(t))$ and $\xi(t)$ satisfying $\|\xi(t)\| \leq R(t)$ and $\rho_i \in [-K, 0]$ such that

$$\begin{aligned}
 &\left\| \int_0^T (\nu(dt) - \xi(t) dt) \right\| < \delta; \\
 &\int_0^T \|\lambda(t) + \xi(t)\| dt < \delta; \\
 &\int_0^T \left\| \left(1 - \sum_i \alpha_{1i} \right) \nu(t) + \mu(t) - \left(1 - \sum_i \alpha_{3i} \right) \left(\int_t^T \xi(\tau) d\tau \right) \right\| dt < \delta; \\
 &\left| \int_0^T \left(\left(1 - \sum_i \alpha_{1i} \right) \langle \nu(t), u_i(t) - y_1(t) \rangle \right. \right. \\
 &\quad \left. \left. - \left(1 - \sum_i \alpha_{3i} \right) \langle \eta(t), u_i(t) - y_3(t) \rangle + g_{i\varepsilon}(t) \right) dt + \rho_i \right| < \delta, \quad i = 1, \dots, k.
 \end{aligned} \tag{11}$$

3. Taking $\delta_m \rightarrow 0$, we shall find sequences of z_{jm} converging to zero as $m \rightarrow \infty$ and of $\nu_m, (\lambda_m(\cdot), \mu_m(\cdot)), \xi_m(\cdot)$ and $\rho_{im} \in [-K, 0], i = 1, \dots, k$ such that

- $\nu_m \in \partial_G(\varphi(z_{1m}))$;
- $(\lambda_m(t), \mu_m(t)) \in \partial_p \tilde{L}(t, x_{2m}(t), y_{2m}(t))$;
- $\|\xi_m(t)\| \leq R(t)$ almost everywhere;
- for every m (11) holds with $\delta = \delta_m, \nu = \nu_m$ etc.

Setting $q_m(t) = -\xi_m(t)$ and $p_m(t) = -v_m(t) - \int_t^T q_m(\tau)d\tau$, we rewrite (11) with some $\gamma_m \rightarrow 0$ and $y_m(\cdot) \rightarrow 0$ as follows:

$$\begin{aligned} &\|p_m(0)\| < \gamma_m; \\ &\int_0^T \|\lambda_m(t) - q_m(t)\|dt < \gamma_m; \\ &\int_0^T \|\mu_m(t) - p_m(t)\|dt < \gamma_m; \\ &\left| \int_0^T (-\langle p_m(t), u_i(t) - y_m(t) \rangle + g_{i\varepsilon}(t))dt + \rho_{im} \right| < \gamma_m, \quad i = 1, \dots, k. \end{aligned} \tag{12}$$

As φ is Lipschitz near zero, total variations of v_m are uniformly bounded and we may assume that v_m weak* converges to some v such that $v \in \partial_G\varphi(0)$. The latter means, in particular, that $v_m(t)$ converge to $v(t)$ at every point of continuity of the latter, that is, everywhere except maybe countably many points. Finally, as all $q_m(\cdot)$ are bounded by the same summable function $R(t)$, this sequence is relatively compact in the weak topology of L^1 . By the Eberlein–Smulian theorem, we can assume that the sequence weak converges to a certain $q(t)$. Hence, the integrals $\int_t^T q_m(s)ds$ uniformly converge to $\int_t^T q(s)ds$, total variations of $p_m(\cdot)$ are uniformly bounded, and $p_m(t)$ converge to

$$p(t) = - \int_t^T q(s)ds - v(t),$$

at every t at which $v(t)$ is continuous.

The second and the third relations in (12) imply that the distance from $(q_m(t), p_m(t))$ to $\partial_p L(t, \cdot)(x_m(t), y_m(t))$ goes to zero for almost every t . On the other hand, by definition of the limiting subdifferential, almost everywhere on $[0, T]$

$$\limsup_{m \rightarrow \infty} \partial_p L(t, \cdot)(x_m(t), y_m(t)) \subset \partial L(t, \cdot)(0, 0).$$

By the Mazur theorem a certain sequence, $\tilde{q}_m(\cdot)$ of convex combinations of $q_m(\cdot)$ converges to $q(\cdot)$ almost everywhere. Hence the distance from $(\tilde{q}_m(t), p_m(t))$ to the set $\{(\tilde{q}, p) : \tilde{q} \in \text{conv} \{q : (q, p) \in \partial L(t, \cdot)(0, 0)\}\}$ goes to zero as $m \rightarrow \infty$. The set $\partial L(t, \cdot)(0, 0)$ is closed and its projection to the q -space is bounded as $L(t, \cdot, y)$ is Lipschitz near zero. Therefore, for any p the set $\text{conv} \{q : (q, p) \in \partial L(t, \cdot)(0, 0)\}$ is also closed and we conclude that $(q(t), p(t))$ belongs to this set for almost every t . The equality $p(0) = 0$ follows from the first relation in (12). This concludes the proof of (4) and (5).

Finally, as sequences (ρ_{im}) are uniformly bounded, we can assume that each of them converges to some $\rho_i \leq 0$. It follows from the last relation in (12) that

$$\int_0^T (g_{i\varepsilon}(t) - \langle p(t), u_i(t) \rangle) dt = -\rho_i \geq 0.$$

This is true for any $\varepsilon > 0$, and when $\varepsilon \rightarrow 0$ the functions $g_{i\varepsilon}(t)$ converge decreasingly to $L(t, 0, u_i(t)) - L(t, 0, 0)$ and we get eventually that

$$\int_0^T (L(t, 0, u_i(t)) - L(t, 0, 0) - \langle p(t), u_i(t) \rangle) dt \geq 0, \quad i = 1, \dots, k. \quad (13)$$

4. We can now conclude the proof. First we note that there is a countable collection \mathcal{U} of bounded measurable selection of $D(\cdot)$ such that for every $u(\cdot) \in \mathcal{U}$ the function $\max_{\|x\| \leq \bar{\varepsilon}} L(t, x, u(t))$ is summable and for almost every t the set $\{u(t) : u(\cdot) \in \mathcal{U}\}$ is dense in $D(t)$.

Indeed, as $D(\cdot)$ is a measurable set-valued mapping, there is a countable collection $\{v_i(\cdot), i = 1, 2, \dots\}$ of measurable selections of $D(\cdot)$ whose values for almost every t form a dense subset of $D(t)$. Take $\varepsilon = \bar{\varepsilon}/2$. Then $\theta_i(t) = \max_{\|x\| \leq \varepsilon} L(t, x, v_i(t))$ is finite-valued by (H_2) . Hence, for every $j = 1, 2, \dots$ there is a subset Δ_{ij} with measure not smaller than $T - j^{-1}$ on which v_i is bounded and $\theta_i(\cdot)$ is summable. Set

$$u_{ij}(t) = \begin{cases} v_i(t), & \text{if } t \in \Delta_{ij}; \\ 0, & \text{otherwise.} \end{cases}$$

Then $\mathcal{U} = \{u_{ij} : i, j = 1, 2, \dots\}$ is a required set.

As we have seen, for any finite collection U of $u(\cdot) \in L^1$ with $u(t) \in D(t)$ a.e. satisfying the conditions of Proposition 2.10 the set $\Lambda(U)$ of triples $(p(\cdot), q(\cdot), v)$ satisfying (4), (5) and such that (13) holds for $u(\cdot) \in U$ is nonempty. As $\|q(t)\| \leq R(t)$ a.e. and the total variations of v are bounded by the Lipschitz constant of φ , the set Λ is compact in the product of the weak*-topologies for functions of bounded variations (that is $p(\cdot)$ and v) and weak L^1 -topology (for $q(\cdot)$). It is also clear that $\Lambda(U') \subset \Lambda(U)$ if $U \subset U'$. Therefore, the intersection of all $\Lambda(U)$, $U \subset \mathcal{U}$ is nonempty. It is also clear that (13) holds for restrictions of $u(\cdot)$ to any measurable subsets of $[0, T]$ (by which we mean the function equals to $u(t)$ on the subset and zero outside).

The standard measurable selection arguments now allow to deduce that this may happen only if for almost every t the inequality $L(t, 0, u) - L(t, 0, 0) - \langle p(t), u(t) \rangle \geq 0$ holds for any $u(\cdot) \in \mathcal{U}$ almost everywhere. This in turn implies (6) since the values of $u(t)$ when $u(\cdot)$ runs through \mathcal{U} and $L(t, \bar{x}(t), \cdot)$ are continuous on $D(t)$. This completes the proof of the theorem.

5 Optimal Control: Proof of Theorem 3.3

As was mentioned in introduction, reduction of the optimal control problem (OC) to the generalized Bolza problem (P) is based on a certain optimality alternative. Here is its statement.

Proposition 5.1 (Optimality alternative [17], Theorem 7,39) *Let X be a complete metric space and M a closed subset of X . Let also f be a function on X that attains a local minimum on M at some $\bar{x} \in M$. If f is Lipschitz near \bar{x} , then for any lsc nonnegative function φ on X equal to zero at \bar{x} the following alternative holds: either there is a $K > 0$ such that $d(x, M) \leq K\varphi(x)$ for all x of a neighborhood of \bar{x} , so that $f + K_0\varphi$ has a local minimum at \bar{x} for some $K_0 > 0$ (not smaller than K times the Lipschitz constant of f), or there is a sequence $(x_n) \rightarrow \bar{x}$ of elements of $X \setminus M$ such that for each n the inequality $\varphi(x) + n^{-1}d(x, x_n) \geq \varphi(x_n)$ holds for all $x \in X$.*

The very fact, which is a simple consequence of Ekeland’s principle, was already used in [9]. The nice feature of the approach based on the optimality alternative is that actually there is no need to verify whether the distance estimate is satisfied or not.

We shall also need the following well-known technical fact. Set for $x(\cdot) \in C[0, T]$

$$\eta(x(\cdot)) = \max_{t \in [0, T]} g(t, x(t)), \quad \Delta(x(\cdot)) = \{t \in [0, T] : g(t, x(t)) = \eta(x(\cdot))\},$$

and let $\bar{\partial}_C g(t, x)$ be the convex hull of the collection of $u \in \mathbb{R}^n$ for which there are sequences $t_m \rightarrow t, x_m \rightarrow x$ and $u_m \rightarrow u$ such that $u_m \in \partial g(t_m, \cdot)(x_m)$.

Proposition 5.2 ([19], Theorem 2.8.2; [21], Example 2.5.2) *Assume (H_4) . Then for any $v \in \partial_C \eta(\bar{x}(\cdot))$ there are a probability measure μ supported on the set $\Delta(\bar{x}(\cdot))$ and a μ -measurable selection of the set-valued mapping $t \rightarrow \bar{\partial}_C g(t, \cdot)(\bar{x}(\cdot))$ such that $v(dt) = \gamma(t)\mu(dt)$.*

Recall that in the statement of the theorem we use a more precise object, namely, $\partial_C^>$. It is an easy matter to see that

$$\partial_C^> g(t, x) = \text{conv} \{ \limsup \bar{\partial}_C g(t_m, x_m) : t_m \rightarrow t, x_m \rightarrow x, g(t_m, x_m) > g(t, x) \}.$$

So let us pass to the proof of the theorem. Again, we can assume that $\bar{x}(t) \equiv 0$. We also assume that $\ell(\bar{x}(0), \bar{x}(T)) = 0$. Then **(OC)** can be equivalently rewritten as

$$\begin{aligned} &\text{minimize } \varphi(x(\cdot)) = \max\{\ell(x(0), x(T)), \max_t g(t, x(t))\}, \\ &\text{s.t. } \dot{x} \in F(t, x), (x(0), x(T)) \in S. \end{aligned}$$

We shall work with this form of **(OC)** in the proof. Finally, let V be a neighborhood of zero in $W^{1,1}$ such that $\varphi(x(\cdot)) \geq 0 = \varphi(0)$ for all feasible $x(\cdot) \in V$.

1. *We shall first prove the theorem under the assumption that $d(y, F(t, \cdot))$ is $\bar{R}(t)$ -Lipschitz for all $y \in \mathbb{R}^n$ (which is the case when $F(t, \cdot)$ is $\bar{R}(t)$ -Lipschitz in the Hausdorff metric). Set*

$$\psi(x(\cdot)) = d((x(0), x(T)), S) + \int_0^T d(\dot{x}(t), F(t, x(t)))dt.$$

Clearly ψ is nonnegative and $\psi(0) = 0$, so we can apply the optimality alternative. Denote by M the feasible set in **(OC)**: $M = \{x(\cdot) \in W^{1,1} : \psi(x(\cdot)) = 0\}$. Then either there is a $K > 0$ such that $d(x(\cdot), M) \leq K\psi(x(\cdot))$ for all $x(\cdot)$ of a neighborhood of zero, so that there is a K_0 such that $\varphi + K_0\psi$ attains an unconditional local minimum at zero (*regular case*), or there are $x_m(\cdot) \in W^{1,1}$, $m = 1, 2, \dots$ converging to zero and such that $\psi(x_m(\cdot)) > 0$ and

$$\begin{aligned} \psi(x(\cdot)) + m^{-1}(\|x(0) - x_m(0)\| + \int_0^T \|\dot{x}(t) - \dot{x}_m(t)\|dt) &> \psi(x_m(\cdot)), \\ \forall x(\cdot) \neq x_m(\cdot). \end{aligned}$$

(*singular case*).

In the regular case, we shall actually consider the problem with slightly different cost functions:

$$\varphi_m(x(\cdot)) = \max\{\ell(x(0), x(T)) + m^{-2}, \max_t g(t, x(t))\}, \quad m = 1, 2, \dots$$

Then $0 = \varphi(0) \leq \varphi(x(\cdot)) \leq \varphi_m(x(\cdot)) \leq \varphi(x(\cdot)) + m^{-2}$ for all feasible $x(\cdot) \in V$. If m is so big that V contains the m^{-1} ball around zero, then by Ekeland’s principle there is a $x_m(\cdot) \in W^{1,1}$ feasible in **(OC)** and such that $\|x_m(\cdot)\|_{1,1} \leq m^{-1}$ and

$$\varphi_m(x(\cdot)) + m^{-1}\|x(\cdot) - x_m(\cdot)\|_{1,1} \geq \varphi_m(x_m(\cdot)) = a_m$$

for all feasible $x(\cdot) \in V$. Clearly, $a_m > 0$ for all sufficiently large m . Otherwise, $\ell(x_m(0), x_m(T))$ would be strictly smaller than zero and $g(t, x_m(t)) \leq 0$ for all t , which contradicts to our assumption that the minimal value of the cost function on V in the original formulation of **(OC)** is zero.

Since $x_m(\cdot) \rightarrow 0$ as $m \rightarrow \infty$, the inequality $d(x(\cdot), M) \leq K\psi(x(\cdot))$ holds for $x(\cdot)$ of a neighborhood of $x_m(\cdot)$. The Lipschitz constants of φ and φ_m coincide; hence, $x_m(\cdot)$ is an unconditional local minimum of $\varphi_m + K_0\psi$.

Summarizing, we conclude that there are $\lambda_0 \in [0, 1]$ ($\lambda_0 = 1$ in the regular case and $\lambda_0 = 0$ otherwise) such that for sufficiently large $m > 0$ the functional

$$\begin{aligned} J_m(x(\cdot)) = \lambda_0\varphi_m(\cdot) + K_0 \left(\int_0^T d(\dot{x}(t), F(t, x(t)))dt + d((x(0), x(T)), S) \right) \\ + m^{-1} \left(\|x(0) - x_m(0)\| + \int_0^T \|\dot{x}(t) - \dot{x}_m(t)\|dt \right) \end{aligned}$$

attains a local minimum at some $x_m(\cdot)$ (feasible in the regular case and not belonging to M in the singular case) with $\|x_m(\cdot)\|_{1,1} < m^{-1}$. In either case, it

is an easy matter to see that J_m satisfies the hypotheses (H_1) and (H_2) and we can apply Theorem 3.1.

Let us start with the *regular case* $\lambda_0 = 1$. Note first that as follows from Propositions 2.3 and 5.2 for any v belonging to the G -subdifferential of φ_m at $x(\cdot)$ there are $\lambda \in [0, 1]$, a positive measure μ with $\mu([0, T]) = 1 - \lambda$ supported on the set Δ_m where $g(t, x_m(t)) = \varphi_m(x_m(\cdot))$, a μ -measurable selection $\gamma(t)$ of the set-valued mapping $\bar{\partial}_C g(t, \cdot)(x(t))$ and a pair of vectors g, h in the subdifferential of $\lambda \ell(\cdot) + d(\cdot, S)$ at $(x(0), x(T))$ such that for any continuous \mathbb{R}^n -valued $u(t)$

$$\int_0^T u(t)v(dt) = \lambda(\langle h, u(T) \rangle + \langle g, u(0) \rangle) + \int_0^T \langle u(t), \gamma(t) \rangle \mu(dt).$$

Thus, applying Theorem 3.1 to J_m and $x_m(\cdot)$ we shall find some $\lambda_m \in [0, 1]$, $g_m, h_m, \mu_m, \gamma_m(\cdot)$, a function $p_m(\cdot)$ of bounded variation, a summable $q_m(\cdot)$ satisfying $\|q_m(t)\| \leq \bar{R}(t)$ a.e. such that $(g_m, h_m) \in \lambda_m \partial(\lambda_0 \ell + K_0 d(\cdot, S))(x_m(0), x_m(T))$, μ_m is a positive measure with $\mu_m([0, T]) = 1 - \lambda_m$ supported on the $\Delta(x_m(\cdot))$ on which $g(t, x_m(t))$ attains maximum, $\gamma_m(\cdot) \in \bar{\partial}_C g(t, \cdot)(x_m(t))$ and the relations

$$\begin{aligned} q_m(t) &\in \text{conv} \{q : (q, p_m(t)) \in \partial L_m(t, \cdot, \cdot)\}; \\ p_m(t) &= -h_m - \int_t^T q_m(s)ds - \int_t^T \gamma_m(s)d\mu_m(s), \quad p_m(0) = g_m; \\ \langle p_m(t), u - \dot{x}_m(t) \rangle &\geq 0, \quad \text{for all } u \in F(t, x_m(t)) \end{aligned} \tag{14}$$

hold for almost every t . Here we have set $L_m(t, x, y) = K_0 d(y, F(t, x)) + m^{-1} \|y - \dot{x}_m(t)\|$.

We may assume that λ_m and (g_m, h_m) converge to some $\lambda \in [0, 1]$ and $(h, g) \in \lambda \partial(\ell + d(\cdot, S))(0, 0)$. Furthermore, since $g(t, x(t))$ is upper semicontinuous, the sets $\Delta(x_m)$ are closed and the excess of $\Delta(x_m)$ over $\Delta(0)$ goes to zero as $x_m(\cdot)$ goes to zero uniformly. Therefore, μ_m weak* converge (along a subsequence of m) to a nonnegative measure μ supported on $\Delta(0)$ and such that $\mu([0, T]) = 1 - \lambda$, whence (i). By (H_4) $\|\gamma_m(t)\| \leq K$, so the sequence $(\gamma_m(\cdot))$ is weakly compact in L^1 which implies (in view of upper semicontinuity of generalized gradients of a Lipschitz function) that for almost any t the value of any weak limit point of this set at t is contained in $\bar{\partial}_C g(t, \cdot)(0)$. If $\mu_m \neq 0$, we necessarily have $\max_t g(t, x_m(t)) = a_m > 0$, whence the appearance of $\partial_C^> g(t, \cdot)(0)$ at the limit.

The sequence $(q_m(\cdot))$ is weakly compact in L^1 (as $q_m(\cdot)$ are bounded by a summable function) and we can assume that it weakly converges to some $q(\cdot)$ in which case $p_m(\cdot)$ converge almost everywhere to

$$p(t) = -h - \int_t^T q(s)ds - \int_t^T \gamma(s)\mu(ds)$$

with $p(0) = g$. We have $p(T) = -h - \gamma(T)\mu(\{T\})$ and (ii) follows from the second relation in (14) (recall that the normal cone to a set is generated by the subdifferential of the distance function to the set at the same point). As $\|\dot{x}_m(t)\| \rightarrow 0$ almost everywhere, the last relation in (14) implies (iv). Finally, a certain sequence of convex combinations of $q_m(\cdot)$ converges to $q(\cdot)$ almost everywhere, and as follows from the first relation in (13), $q(t) \in \text{conv} \{q : (q, p(t)) \in \partial L(t, \cdot, \cdot)(0, 0)\}$, with $L(t, x, y) = K_0 d(y, F(t, x))$, whence (iii). This completes the proof in the regular case.

In the *singular case*, when $\lambda_0 = 0$, the same arguments work in substantially simpler situation as the off-integral term reduces to $d((x(0), x(t)), S) + m^{-1} \|x(0) - x_m(0)\|$. The only difference is the proof of (i). In the singular case, $\psi(x_m(\cdot)) > 0$. This means that for any m either $d((x_m(0), x_m(T)), S) > 0$ in which case $\max\{\|g_m\|, \|h_m\|\} = 1$, or $d(\dot{x}_m(t), F(t, x_m(t))) > 0$ on the set of positive measure which means that $\|p_m(t)\| \geq 1 - m^{-1}$ at least at one point. Thus, in either case $\|p_m(t)\| \geq 1 - m^{-1}$ at least at one point. On the other hand, as $\mu_m = 0$ in the singular case, $p_m(\cdot)$ are continuous functions converging uniformly, so that $p(\cdot) \neq 0$. This completes the proof of the theorem for the case when $d(y, F(t, \cdot))$ is $\bar{R}(t)$ -Lipschitz for any y .

- Let us turn to the proof under the stated conditions. First we apply Proposition 2.7 to $F = F(t, \cdot)$, $C = \{0\}$, $r = \bar{r}(t)$, $R = \bar{R}(t)$, $\varepsilon = \bar{\varepsilon}$ and $\eta \in [\bar{\eta}, 1)$. Let $\Gamma_0(t, x)$ be the corresponding mapping into $\mathbb{R} \times \mathbb{R}^n$. It is obviously measurable.

By (H5), $\bar{r}(t)$ is bounded away from zero, that is, there is an $\bar{r} > 0$ such that $\bar{r}(t) \geq \bar{r}$ for almost all t . Fix some positive $r < \bar{r}$, $R > 0$, $\eta \in [\bar{\eta}, 1]$ and $\varepsilon \in (0, \bar{\varepsilon})$ and let $C(t)$ be the set of all $u \in F(t, 0)$ such that (9') holds with chosen r, R, η and ε . Clearly, $C(t)$ is a closed set (that in principle can be empty). Again, it is not a difficult matter to verify that $C(\cdot)$ is a measurable mapping. (Consider the functions

$$g(t, u) = \sup\{\text{ex}(F(x) \cap B(u, r), F(x' + R(\|x' - x\|B))) : x, x' \in \varepsilon B\};$$

$$h(t, u) = \max\{d(F(t, x), B((u, (1 - \eta)r)))\}$$

(where $\text{ex}(A, B)$ is the excess of A over B and $d(A, B)$ is the distance between A and B) and the sets of $u \in F(t, 0)$ where both are equal zero.)

We shall slightly modify $C(t)$: let $C'(t) = \{0\}$ for t in a set Q_ε of small positive measure, say smaller than ε , and otherwise $C'(t) = C(t)$. Let $\Gamma'(t)$ be the mapping into $\mathbb{R} \times \mathbb{R}^n$ corresponding to $C'(t)$ and the chosen r, R, ε and η according to Proposition 2.7. Set $\Gamma(t, x) = \Gamma_0(t, x) \cup \Gamma'(t, x)$ and consider the problem

$$\begin{aligned} &\text{minimize } \varphi(x(\cdot)), \\ &\text{s.t. } (\dot{\xi}, \dot{x}) \in \Gamma(t, x), (x(0), x(T)) \in S, \xi(0) = 0, \xi(T) = T. \end{aligned}$$

Then $\xi(t) = t$ for any feasible pair $(\xi(\cdot), x(\cdot))$, which means that $x(\cdot)$ is feasible in (OC). Hence, $(t, 0)$ is a local minimum in the problem and we can apply the result obtained in the first part of the proof to get a necessary optimality condition. It

follows that there are $\lambda \geq 0$, a nonnegative measure μ supported on the set $\Delta(0)$, a μ -measurable $\gamma(t) \in \partial_{\mathcal{C}}^{\geq} g(t, 0)$, \mathbb{R}^{n+1} -valued function $(\kappa(t), p(t))$ of bounded variation and summable $(\omega(t), q(t))$ such that

$$\begin{aligned} \lambda + \|(\kappa(\cdot), p(\cdot))\| + \mu([0, T]) &= 1; \\ (p(0), -p(T) - \gamma(T)\mu(\{T\})) &\in \lambda \partial \ell(0, 0) + N(S, (0, 0)); \end{aligned} \tag{15}$$

and almost everywhere on $[0, T]$

$$\begin{aligned} \dot{\kappa}(t) &= \omega(t), \quad dp(t) = q(t)dt + \gamma(t)\mu(dt); \\ (\omega(t), q(t)) &\in \text{conv} \{(\omega, q) : ((\omega, q), (\kappa(t), p(t))) \in \partial G(t, \cdot, \cdot)((t, 0), (1, 0))\}; \\ \omega(t)(\chi - 1) + \langle p(t), u \rangle &\leq 0, \quad \forall (\chi, u) \in \Gamma(t, 0), \end{aligned} \tag{16}$$

where $G(t, (\xi, x), (\sigma, y)) = K_0 d((\sigma, y), \Gamma(t, (\xi, x)))$ and K_0 is a certain positive number. Since Γ does not depend on ξ , we deduce that $\omega(t) \equiv 0$, and therefore, $\kappa(t) = \text{const} = \kappa$. We claim that actually $\kappa(t) \equiv 0$ and the second relation in (16) reduces to

$$q(t) \in \text{conv} \{q : (q, p(t)) \in K \partial L(t, \cdot, \cdot)(0, 0)\}, \tag{17}$$

where we have set $L(t, x, y) = d(y, F(t, x))$. This would prove the theorem for $U(t)$ replaced by the union of $\bar{r}(t)B$ and $C(t)$.

Let $t \in Q_\varepsilon$ and $((0, q), (\kappa, p(t))) \in \partial G(t, \cdot, \cdot)((t, 0), (1, 0))$. Recall that $\Gamma(t, \cdot) = \Gamma_0(t, \cdot)$ for such t . Choose further small positive δ and θ to ensure that $(1 - \eta)\bar{r} > 3\delta$ and $\theta \bar{R}(t) < \delta$. The latter implies by (H5) that $d(0, F(t, x)) < \delta$ if $\|x\| < \theta$.

We have $d((\sigma, y), \Gamma_0(t, x)) = \inf\{\|\sigma - v\| + \|y - vz\| : z \in F(t, x), \|z\| \leq (1 - \sigma v)\bar{r}(t)\}$. However, if $1 - \sigma < \delta$, $\|y\| < \delta$, $\|x\| < \theta$, then the inequality $\|z\| \leq (1 - \sigma v)\bar{r}(t)$ is automatically satisfied for (v, z) realizing the infimum. Hence, $d((\sigma, y), \Gamma_0(t, x)) = \inf\{\|\sigma - v\| + \|y - vz\| : z \in F(t, x)\}$. On the other hand, we can be sure that $d((y/v), F(t, x)) < 1$ for (v, y) close to $(1, 0)$. Therefore, the infimum is attained with $v = \sigma$ and is equal to $vd((y/v), F(t, x)) = vL(t, x, y/v)$.

Set $g_t(v, x, y) = vL(t, x, y/v)$. We can view g_t as a composition $h_t \circ A$, where $h_t(v, x, y) = vL(t, x, y)$ and $A(v, x, y) = (v, x, y/v)$ is a smooth map. Applying Proposition 2.5 (with a reference to Remark 2.2), we get the equality $\partial g_t(1, 0, 0) = (0, \partial L(t, \cdot, \cdot))(0, 0)$ which proves the claim.

Denote by Λ , the collection of Lagrange multipliers $\lambda, p(\cdot), q(\cdot), \nu$ (where $\nu(dt) = \gamma(t)\mu(dt)$) satisfying (15) (with $\kappa(\cdot) = 0$), (17). By Proposition 2.9, $d|p|(t) \leq \bar{R}(t)|p|(t) + d|\nu|(t)$, where $|p|$ and $|\nu|$ stand for variations of p and ν . The total variation of ν is bounded by the Lipschitz constant of φ ; hence, by Grownwall’s lemma the total variation of $p(\cdot)$ is also bounded and (again by Proposition 2.9) $\|q(t)\| \leq \bar{R}(t)|p|([0, T])$. This means that Λ is a compact set if $p(\cdot)$ and ν are considered with the weak*-topology of measures and $q(\cdot)$ with the weak topology of L^1 .

Let now $r_i \rightarrow 0, \eta_i \rightarrow 0$ (both decreasingly), $R_i \rightarrow \infty$ (increasingly) and $\varepsilon_i \rightarrow 0$ (decreasingly) satisfy $\varepsilon_i R_i < (1 - \eta_i)r_i$, and let $C'_i(t)$ be the corresponding subsets of

$F(t, 0)$. If Q_{ε_i} are chosen to guarantee that $Q_{\varepsilon_{i+1}} \subset Q_{\varepsilon_i}$ (which is of course possible), then $C_i(t) \subset C_{i+1}(t)$ and every $u \in U(t)$ belongs to $C'_i(t)$ if i is sufficiently large. As we have seen, for any i the set of Λ_i of multipliers guaranteeing that the conclusion of the theorem holds with $U(t)$ replaced by $(\bar{F}(t)B) \cup C_i(t)$ in (iv) is nonempty. It is obvious that $\Lambda_{i+1} \subset \Lambda_i$, and all Λ_i are compact, $\cap \Lambda_i \neq \emptyset$. This completes the proof of the theorem.

Remark 5.1 The only purpose of introducing functionals φ_m in the regular case was to get more precise information about location of values of $\gamma(\cdot)$. In case when we know a priori that $\bar{\partial}_C g(t, \cdot)$ and $\partial_C^> g(t, \cdot)$ coincide (e.g., when $g(t, \cdot)$ is differentiable and the derivative is continuous in both variables), the proof will be noticeably shorter.

6 An Open Problem

If we know a priori that **(OC)** is regular, that is, the distance to the feasible set from any $x(\cdot)$ of a $W^{1,1}$ -neighborhood of $\bar{x}(\cdot)$ is majorized by $K(\int_0^T d(x(t), F(t, x(t)))dt + d((x(0), x(T)), S))$, then the conclusion of the theorem can be significantly strengthened. Indeed, regularity would allow us to reduce the problem to unconstrained minimization of

$$\varphi(x(\cdot)) + K_0 \left(\int_0^T d(x(t), F(t, x(t)))dt + d((x(0), x(T)), S) \right).$$

Applying Theorem 3.1 to this functional, we would deduce that the conclusion of the theorem is valid with $U(t)$ replaced by the set of $u \in F(t, \bar{x}(t))$ such that the function $d(u, F(t, \cdot))$ is continuous in the $\bar{\varepsilon}$ -neighborhood of $\bar{x}(t)$. The question is whether such a replacement can be justified also in the singular case. Positive answer to the question would allow to extend (iv) to those $u \in F(t, \bar{x}(t))$ for which the function $d(u, F(t, \cdot))$ is continuous near $\bar{x}(t)$.

Similar extension is possible for optimal control problems in the classical Pontryagin form when the right side of the equation is continuous with respect to the state variable (see, e.g., [22] and references therein).

7 Conclusions

As is stressed in the introduction, the purpose of the paper has been not just to prove new and stronger necessary optimality conditions but also to show that how a heavily constrained optimal control problem can be equivalently reduced to an unconstrained Bolza problem with simply structured integrand and off-integral term. It is to be once again emphasized that such a reduction can be performed and effectively used for further analysis only in the framework of modern variational analysis that allows to work with nonsmoothness and set-valuedness. I believe that the approach has a strong potential and can be effectively used far beyond first-order optimality conditions, in particular, to study second-order conditions for a strong minimum in optimal control and the Hamilton–Jacobi theory.

Acknowledgements I wish to express my gratitude to the anonymous referee for helpful remarks.

References

1. Clarke, F.H.: Necessary conditions in dynamic optimization. Mem. AMS **816**, 113 (2005)
2. Vinter, R.B.: Optimal Control. Birkhauser, Basel (2000)
3. Ioffe, A.D.: Necessary and sufficient conditions for a local minimum 1. Reduction theorem and first order conditions. SIAM J. Control Optim. **17**, 245–251 (1979)
4. Ioffe, A.D., Tikhomirov, V.M.: Theory of Extremal Problems, Nauka, Moscow 1974 (**in Russian**) (English transl: North Holland) (1979)
5. Rockafellar, R.T.: Existence theorems for general control problems of Bolza and Lagrange. Adv. Math. **15**, 312–333 (1975)
6. Clarke, F.H.: The generalized problem of Bolza. SIAM J. Control Optim. **14**, 682–699 (1976)
7. Clarke, F.H.: The maximum principle under minimal hypotheses. SIAM J. Control Optim. **14**, 1078–1091 (1976)
8. Loewen, P.D.: Optimal control via nonsmooth analysis. In: CRM Proceedings and Lecture Notes, vol. 2. AMS (1993)
9. Ioffe, A.D.: Euler–Lagrange and Hamiltonian formalisms in dynamic optimization. Trans. Am. Math. Soc. **349**, 2871–2900 (1997)
10. Ioffe, A.D., Rockafellar, R.T.: The Euler and Weierstrass conditions for nonsmooth variational problems. Calc. Var. PDEs **4**, 59–87 (1996)
11. Smirnov, G.V.: Discrete approximations and optimal solutions to differential inclusions, *Kibernetika* (Kiev) 1991, (1), 76–79 (**in Russian**); [English transl.: *Cybernetics* **27**(1), 101–107 (1991)]
12. Loewen, P.D., Rockafellar, R.T.: Optimal control of unbounded differential inclusions. SIAM J. Control Optim. **32**, 442–470 (1994)
13. Mordukhovich, B.S.: Optimization and finite difference approximations of nonconvex differential inclusions with free time. In: Mordukhovich, B.S., Sussmann, H.J. (eds.) *Nonsmooth Analysis and Geometric Methods in Deterministic Optimal Control*, pp. 153–202. Springer, Berlin (1996)
14. Vinter, R.B., Zheng, H.: The extended Euler–Lagrange conditions in nonconvex variational problems. SIAM J. Control Optim. **35**, 56–77 (1997)
15. Rockafellar, R.T., Wets, R.J.B.: *Variational Analysis*. Springer, Berlin (1998)
16. Clarke, F.H., Ledyaev, Yu.S., Stern, R.J., Wolenski, P.R.: *Nonsmooth Analysis and Control Theory*. Springer, Berlin (1998)
17. Ioffe, A.D.: *Variational Analysis of Regular Mappings*. Springer, Berlin (2017)
18. Penot, J.-P.: *Calculus Without Derivatives*, Graduate Texts in Mathematics, vol. 266. Springer, Berlin (2012)
19. Clarke, F.H.: *Optimization and Nonsmooth Analysis*. Wiley-Interscience, New York (1983)
20. Bogolyubov, N.N.: Sur quelques méthodes nouvelles dans le calcul des variations. Ann. Math. Pure Appl. Ser. 4 **7**, 249–271 (1930)
21. Ioffe, A.D.: Necessary conditions in nonsmooth optimization. Math. Oper. Res. **9**, 159–189 (1984)
22. Arutyunov, A.V., Vinter, R.B.: A simple finite approximations proof of the Pontryagin maximum principle, under deduced differentiability Hypotheses. Set Valued Anal. **12**, 5–24 (2004)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.