IIASA
International Institute for Applied Systems Analysis • A-2361 Laxenburg • Austria
Tel: +43 2236 807 • Fax: +43 2236 71313 • E-mail: info@iiasa.ac.at • Web: www.iiasa.ac.at

INTERIM REPORT IR-97-030 / June

Optimal Control of Nonconvex Differential Inclusions

Boris S. Mordukhovich (boris@math.wayne.edu)

Approved by
Arkadii Kryazhimskii (kryazhim@iiasa.ac.at)
Senior Research Scholar, *Dynamic Systems Project*

Contents

Abstract

Optimization problems for discrete and differential inclusions have many important applications and generalize both standard and nonstandard models in optimal control for open-loop and closed-loop control systems. In this paper we consider optimal control problems for dynamic systems governed by such inclusions with general endpoint constraints. We provide a variational analysis of differential inclusions based on their finite difference approximations and recent results in nonsmooth analysis. Using these techniques, we obtain refined necessary optimality conditions for nonconvex-valued discrete and differential inclusions in a general setting. These conditions are expressed in terms of robust nonconvex generalized derivatives for nonsmooth mappings and multifunctions. We also provide a brief survey of recent results in this direction.

Key words and phrases: optimal control, nonconvex differential inclusion, discrete approximation, nonsmooth analysis, coderivative, Euler-Lagrange formalism.

Mathematical Subject Classifications (1991): 49K24, 49J52, 49M25.

About the Author

Boris S. Mordukhovich is a Full Professor at the Department of Mathematics, Wayne State University, Detroit, Michigan 48202, U.S.A., and member of the Editorial Board of SIAM Journal on Control and Optimization.

Optimal Control of Nonconvex Differential Inclusions

Boris S. Mordukhovich (boris@math.wayne.edu)

1 Introduction

The main intention of this paper[1] is devoted to optimal control problems for differential inclusions whose right-hand side may be *nonconvex-valued.* Such a case is typical for nonlinear control systems and is particularly important for applications. As a basic model, we consider the following problem (P):

$$\text{minimize} \;\; I(x(\cdot)) := \varphi_0(x(a), x(b)) \tag{1}$$

over absolutely continuous trajectories $x : [a, b] \to \mathbf{R}^n$ of the differential inclusion

$$\dot{x}(t) \in F(x(t), t) \;\; \text{a.e.} \;\; t \in [a, b] \tag{2}$$

on the fixed time interval $[a, b]$ under the general endpoint constraints:

$$\varphi_i(x(a), x(b)) \le 0 \;\; \text{for} \;\; i = 1, 2, \dots, q; \tag{3}$$

$$\varphi_i(x(a), x(b)) = 0 \;\; \text{for} \;\; i = q+1, \; q+2, \dots, q+r; \tag{4}$$

$$(x(a), x(b)) \in \Omega \subset \mathbf{R}^{2n}. \tag{5}$$

It is well known that many other optimization problems for differential inclusions can be reduced to the form (P) (e.g., problems with integral cost functionals and isoperimetric constraints, problems over non-fixed time intervals, etc.). Note also that the differential inclusion (1.2) covers open-loop control systems of the usual form

$$\dot{x} = f(x, u, t) \;\; \text{with} \;\; u(t) \in U(t) \tag{6}$$

and, moreover, this model allows one to consider closed-loop control systems with $F(x, t) = f(x, U(x, t), t)$.

The mainstream in studying optimal control problems for differential inclusions consists of obtaining necessary conditions for optimal solutions. There are different approaches and various results in this area (see, e.g., [1, 5, 6, 8, 10, 12, 14, 18, 22, 32, 33, 35] and references therein). Most of these approaches treat (P) as an infinite dimensional variational problem and employ one or another technique in *nonsmooth analysis.* We are not going to present here a detailed survey of all the achievements in necessary optimality conditions for differential inclusions referring the reader to the papers mentioned above.

[1]This research was partially supported by the National Science Foundation grant DMS-9404128, the USA-Israel BSF grant 94-00237, and the NATO contract CRG-950360.

In what follows we concentrate on *intrinsic* conditions expressed in terms of generalized derivatives of the initial data when the admissible velocity sets $F(x,t)$ are *bounded and Lipschitzian* in x.

Up to the latest time, intrinsic necessary optimality conditions for differential inclusions were obtained under the *convexity* of $F(x,t)$, unless strong regularity assumptions on endpoint constraints were imposed. Note that the convexity hypothesis is rather restrictive and does not hold in many important applications. For example, in the case of control systems (1.6) this assumption is close to the linearity of f with respect to u and the convexity of the control set $U(t)$. Note also that the Pontryagin maximum principle for systems (1.6) is proved with no convexity assumptions. Moreover, when the sets $F(x,t)$ are convex and Lipschitz continuous with respect to x, the differential inclusion (1.2) can be represented, due to Lojasiewicz'es parametrization theorem, in the form of control system (1.6) with a Lipschitzian function $f(\cdot,u,t)$; see, e.g., [1, 35] for more details and references.

In contrast to the parametrized control systems (1.6), the case of differential inclusions admits different forms of the *adjoint inclusion* in necessary optimality conditions. There are two well-known Clarke's independent conditions for problems like (P) with Lipschitzian data: the Euler-Lagrange inclusion [3] and the Hamiltonian inclusion [4, 5] proved under the convexity hypothesis. The first of them is expressed in terms of Clarke's normal cone to the graph of F and the second one is represented in terms of his generalized gradient of the Hamiltonian associated with (1.2).

Another version of the Euler-Lagrange inclusion for problem (P) under the convexity of $F(x,t)$ was established by Mordukhovich [17] (see also [18, 20]) in terms of the *coderivative* $D^\star F$ associated with the *nonconvex* normal cone to the graph of F; see Section 3. The latter construction, that may often be substantially smaller than the Clarke normal cone, was first used in the earlier work [16] to obtain refined transversality conditions in nonsmooth optimal control problems. It has been recently proved by Rockafellar [29] (in one direction) and Ioffe [10] (in the other one) that the Euler-Lagrange inclusion in the form of [17] is *equivalent* to Clarke's Hamiltonian inclusion for convex-valued problems (P).

The main concern of this paper is the Euler-Lagrange inclusion in the *refined form*

$$\dot{p}(t) \in \text{co}\ D_x^\star F(\bar{x}(t), \dot{\bar{x}}(t), t)(-p(t)) \ \text{ a.e.} \tag{7}$$

and its discrete counterpart. In (1.7), $D_x^\star F$ means the coderivative [17] of F in x at the point $(\bar{x}, \dot{\bar{x}}(t), t)$ and "co" stands for convex hull. Inclusion (1.7) (as a necessary condition for optimal solutions to (P)) coincides with the one in [17] when the set $F(\bar{x}(t),t)$ is *strictly convex* a.e. in $[a,b]$. Later Smirnov [32] and Loewen and Rockafellar [14] independently extended this result to general classes of convex-valued differential inclusions by different methods involving certain strict convexification procedures.

Furthermore, Rockafellar [30] recently established that (1.7) is *equivalent* in the case of convex $F(x,t)$ to a new Hamiltonian-type condition of the form

$$\dot{p}(t) \in \text{co}\,\{w|\ (-w, \dot{\bar{x}}(t)) \in \partial H(\bar{x}(t), p(t), t)\} \ \text{ a.e.} \tag{8}$$

where ∂H is the nonconvex subdifferential (associated with the coderivative in (1.7); see Section 3) of the Hamiltonian

$$H(x,p,t) := \sup\{\langle p, v\rangle |\ v \in F(x,t)\} \tag{9}$$

with respect to (x,p). The latter Hamiltonian condition obviously sharpens that of Clarke [5], since (1.8) involves the convexification with respect to only one component.

Note that Rockafellar obtained the equivalence result under the convexity of $F(x,t)$ and certain continuity assumptions that are implied by Lipschitzian behavior. Recently Ioffe [10] established the implication (1.7)$\Longrightarrow$(1.8) without the additional continuity assumption in [30] but still under the convexity of $F(x,t)$.

Thus, for the case of convex-valued differential inclusions there are two equivalent adjoint inclusions (1.7) and (1.8) forming the best intrinsic necessary conditions for strong minimizers. We refer the reader to the recent paper by Loewen and Rockafellar [15] that contains state-of-the-art results, based on (1.7) and (1.8), for more general dynamic optimization problems involving convexity with respect to velocity variables.

The case of *nonconvex* differential inclusions turns out to be more complicated. There is still unsolved Clarke's long-standing conjecture about the validity of his Hamiltonian inclusion as a necessary condition for strong minimizers in variational problems with general endpoint constraints; see [6], p. 71. As for Clarke's form of the Euler-Lagrange inclusion, its necessity in (P) follows from controllability results of Kaśkosz and Lojasiewicz [12] for nonconvex differential inclusions. Let us also mention the recent paper of Zhu [35] who established a "non-intrinsic," parametrized form of necessary optimality conditions for nonconvex problems following the road mapped by Warga [34].

The refined form (1.7) of the Euler-Lagrange condition for nonconvex differential inclusions was first obtained by Mordukhovich [22] based on discrete approximations and appropriate tools of nonsmooth analysis in finite dimensions. Associated results involving (1.7) were also established for nonconvex problems with Bolza-type functionals, boundary trajectories, "intermediate" (in $W^{1,p}$, $1 \le p < \infty$) local minima, free time, etc.; see [22, 23, 26] for more details.

Let us observe that the Euler-Lagrange inclusion (1.7) automatically implies the Weierstrass-Pontryagin maximum condition

$$\langle p(t), \dot{\bar{x}}(t)\rangle = H(\bar{x}(t), p(t), t) \ \text{ a.e.} \tag{10}$$

when the sets $F(x,t)$ are convex around $\bar{x}(\cdot)$. This follows directly from properties of the coderivative for convex-valued multifunctions; see Section 3. It is no longer true in the nonconvex case.

It has been indicated in [22] (see the proof of Thm. 7.1 and Rem. 7.6) that the Euler-Lagrange inclusion (1.7) is accomplished by the the maximum condition (1.10) for nonconvex problems (P) provided that this inclusion is supplemented by the classical Weierstrass inequality for an *unconstrained* nonconvex problem of Bolza; cf. also [23]. The latter result has been recently established by Ioffe and Rockafellar [11] in the general framework of nonconvex and nonsmooth Bolza problems with no dynamic constraints (finite Lagrangian). The proof in [11] is based on tools of infinite dimensional nonsmooth analysis for integral functionals.

An alternative proof of the same result was later proposed by Vinter and Zheng [33]. Their proof is simpler than that in [11] providing a reduction of the nonconvex variational problem of Bolza to an optimal control problem with *smooth dynamics* and *nonsmooth endpoint constraints* in the geometric form (1.5). This is in fact the case treated originally in [16] where the Pontryagin maximum principle was obtained with transversality conditions in terms of the nonconvex normal cone; see below. It is worth mentioning that the proof of the latter result in [16] is based on reducing the nonsmooth problem under consideration to a sequence of *smooth* optimal control problems with *free endpoints* where the Pontryagin maximum principle admits a quite elementary derivation.

Furthermore, general dynamic optimization problems for nonconvex differential inclusions can be reduced, in turn, to minimizing nonsmooth and nonconvex Bolza functionals with no dynamics constraints. There are several schemes of such a reduction (approximation) involving Ekeland's variational principle; see, e.g., [4, 10, 22, 33]. In this way, employing nonconvex subdifferential calculus, one can derive the refined Euler-Lagrange inclusion (1.7) supplemented by the maximum and transversality conditions for optimal control problems like (P) from the corresponding nonconvex versions of the Euler-Lagrange, Weierstrass, and transversality conditions for unconstrained variational problems of Bolza. We refer the reader to [10, 22, 23, 33] for more details and exact assumptions under which the conditions discussed are necessary for optimal solutions to nonconvex differential inclusions.

In this paper we consider basic constructions of the *discrete approximation method* to study optimization problems for differential inclusions. Such a direct method, based on finite difference replacements of the derivative in (1.2), goes back to Euler (1744) who employed it to derive the classical Euler-Lagrange equation in the calculus of variations. It is well known that finite difference and related approximation methods provide powerful algorithmic tools for numerical solutions of infinite dimensional variational problems; see, e.g., the recent survey in Dontchev and Lempio [7]. The present paper mainly explores *theoretical* advantages of discrete approximations that allow us to obtain qualitative results for infinite dimensional optimization problems via appropriate finite dimensional approximations. To furnish this one should first focus on constructing correct discrete approximations that ensure proper *convergence* results for optimal solutions. We consider these questions in Section 2.

Having in hand well-posed discrete approximations with appropriate convergence results, one can study optimal solutions for corresponding finite dimensional problems and then obtain optimality conditions for the original problem in infinite dimensions by passing to the limit in approximations. For the case of variational problems involving differential inclusions (even with smooth objectives and no endpoint constraints), this procedure requires the usage of appropriate tools of *nonsmooth analysis*. Main complications come from the dynamic constraint (1.2) and its finite difference approximations that lead to an increasing number of (nonsmooth and nonconvex) geometric constraints in corresponding finite dimensional problems. In Section 3 we present basic results of generalized differentiation that allow us to handle such situations and support limiting procedures.

The final Section 4 deals with necessary optimality conditions for nonconvex discrete and differential inclusions. We also discuss some results and open questions related to the *approximate maximum principle* in parametric families of discrete approximations that take an intermediate position between optimal control systems with continuous and discrete time.

Our notation is basically standard. Recall that cone $\Omega := \{\alpha x : \alpha > 0,\ x \in \Omega\}$ means the conic hull of Ω; the set B is always the unit closed ball of the space in question; the set

$$\mathrm{Limsup}_{x\to\bar{x}} F(x) := \{y \in \mathbf{R}^m |\ \exists \text{ sequences } x_k \to \bar{x},\ y_k \to y \text{ with } y_k \in F(x_k)\}$$

is the Kuratowski-Painlevé upper limit of the multifunction $F : \mathbf{R}^n \Rightarrow \mathbf{R}^m$; and Ker $F := \{x \in \mathbf{R}^n |\ 0 \in F(x)\}$.

2 Discrete Approximation

First let us construct a finite difference (discrete) approximation of the differential inclusion (1.2) using the replacement of the derivative by the *Euler finite difference*

$$\dot{x}(t) \approx [x(t+h) - x(t)]/h \ \text{ as } \ h \downarrow 0.$$

For any natural number $N = 1, 2, \ldots$ we consider a uniform grid $T_N := \{a, a + h_N, \ldots, b - h_N\}$ with the stepsize $h_N := (b - a)/N$. We associate with (1.2) the following sequence of discrete inclusions:

$$x_N(t + h_N) \in x_N(t) + h_N F(x_N(t), t) \text{ for } t \in T_N, \quad N = 1, 2, \ldots. \tag{11}$$

For any $t \in [a, b]$ we denote by t_N a point of the grid T_N nearest to t from the left. Let us consider piecewise linear *extensions* of discrete trajectories

$$x_N(t) := x_N(t_N) + h_N^{-1}[x_N(t + h_N) - x_N(t)](t - t_N), \quad t \in [a, b],$$

with the piecewise constant velocities $\dot{x}_N(t),\ t \in [a, b] \setminus T_N,\ N = 1, 2, \ldots$.

Could one approximate and in which sense trajectories of the differential inclusion (1.2) by extended trajectories of its discrete counterparts (2.1)? This question is basic for both theoretical and numerical aspects of discrete approximations. Conventional results in this direction relate to the uniform (C-space) convergence of discrete approximations; see, e.g., [7]. The latter convergence is *not* sufficient to study nonconvex differential inclusions since it corresponds to the weak (in L^1) convergence of velocities and requires the convexification of the velocity sets $F(x, t)$. What we actually need is the *pointwise* (strong) convergence of the extended discrete velocities $\dot{x}_N(t)$ a.e. in $[a, b]$. The most recent result on such a strong approximation is proved in [22] under the following hypothesis:

(H1) There is an open set $W \subset \mathbf{R}^n$ where the multifunction $F(x, t)$ is compact-valued, Lipschitz continuous in x, and Hausdorff continuous in t a.e. in $[a, b]$.

2.1. Theorem. *Let $x(\cdot)$ be a trajectory of the differential inclusion* (1.2) *with $x(t) \in W$ in $[a, b]$. Then there is a sequence of extended trajectories $x_N(\cdot)$ for discrete inclusions* (2.1) *such that $x_N(a) = x(a)$ and $\dot{x}_N(t) \to \dot{x}(t)$ a.e. in $[a, b]$.*

The proof of this theorem is based on a *proximal algorithm* involving projections of velocities instead of projections of states; cf. [18, 22, 32] and their references. In this way we provide effective *error estimates* to the rate of strong approximation; see Section 3 in [22] for more details.

Note that Theorem 2.1 deals with discrete approximations of *any* trajectory for the differential inclusion (1.2) under mild assumptions on its initial data. Next we consider approximations for the dynamic *optimization* problems (P) involving (1.2) and the endpoint constraints (1.3)–(1.5). The point here is to approximate not only the dynamic constraint (1.2) but also the endpoint constraints and, moreover, *match* these approximations in order to obtain appropriate convergence results for optimal solutions. In this way we consider the following sequence (P_N) of discrete approximations to the original problem (P):

$$\text{minimize } I_N(x_N(\cdot)) := \varphi_0(x_N(a), x_N(b)) \tag{12}$$

over all trajectories of the difference inclusions (1.2) under the perturbed endpoint constraints

$$\varphi_i(x_N(a), x_N(b)) \le \gamma_{iN} \quad \text{for } i = 1, 2, \ldots, q; \tag{13}$$

$$-\eta_{iN} \le \varphi_i(x_N(a), x_N(b)) \le \eta_{iN} \quad \text{for } i = q + 1,\ q + 2, \ldots, q + r; \tag{14}$$

$$(x_N(a), x_N(b)) \in [\Omega]_{\rho_N} \tag{15}$$

where $[\cdot]_\rho$ stands for the Hausdorff ρ-neighborhood of the set, and $\gamma_{iN} \to 0,\ \eta_{iN} \downarrow 0,\ \rho_N \to 0$ as $N \to \infty$.

To formulate basic results about convegence of discrete approximations $(P_N) \to (P)$ we first look at an internal property of the continuous-time problem (P) called *relaxation stability.* Along with (P) we consider its *relaxation* (R) that consists of minimizing the cost functional (1.1) over absolutely continuous trajectories $x(t),\ a \le t \le b$, for the *convexified* differential inclusion

$$\dot{x}(t) \in \text{co}\ F(x(t), t) \ \text{ a.e. } \ t \in [a, b] \tag{16}$$

under the same endpoint constraints (1.3)–(1.5). Denoting by $\inf(P)$ and $\inf(R)$ the infimum of the cost functional in problems (P) and (R) respectively, one always has

$$\inf(R) \le \inf(P). \tag{17}$$

We say that problem (P) is *stable with respect to relaxation* if (2.7) holds as equality.

It is well known that relaxation stability is a common property of continuous-time control systems that spreads far beyond systems with the convex admissible velocity sets $F(x, t)$. This property is connected with the so-called "hidden convexity" of nonconvex differential systems and is fulfilled in many important situations (see, e.g., [5, 22, 34] and references therein). The key fact here is that, under standard assumptions (in particular, for the case of locally Lipschitzian multifunctions $F(\cdot, t)$), any trajectory of the relaxed system (2.6) can be *uniformly approximated* by trajectories of the initial inclusion (1.2) starting from the same point; see, e.g., [5], p.117. This fact implies the relaxation stability of any problem (P) with (trivial) endpoint constraints (1.3)–(1.5) localized at either the left-hand end $t = a$ or at the right-hand end $t = b$ of admissible trajectories $x(t)$.

Note that results in this vein go back to Bogoljubov's theorem [2] that ensures the property

$$\inf \int_a^b g(x(t), \dot{x}(t), t)dt = \int_a^b \text{co}\ g(x(t), \dot{x}(t), t)dt \tag{18}$$

in the classical Bolza problem with arbitrary endpoint constraints where co $g(x, v, t)$ stands for the convexification of the function $v \to g(x, v, t)$. In (2.8) the integrand g is continuous in (x, v) and may be measurable in t; see Sec. 2 in [22] for more details.

In the case of general differential inclusions with nontrivial endpoint constraints, the relaxation stability of (P) is ensured by the *calmness* property of this problem [5] that holds for "almost all" boundary conditions. Let us also mention results of Warga [34] who proved the relaxation stability of control problems admitting *normal* necessary optimality conditions.

For special classes of differential inclusions, the relaxation stability holds for *arbitrary* endpoint constraints with *no* calmness or normality assumptions. In particular, let

$$\dot{x}(t) \in F_1(t)x(t) + F_2(t) \ \text{ a.e.}$$

where both multifunctions F_1 and F_2 are integrable in $[a, b]$ and, in addition, F_1 is convex-valued while F_2 is not. Then the relaxation stability of problem (P) can be proved for any endpoint constraints (1.3)–(1.5) using Lyapunov-Aumann's theorem about set-valued integrals; cf. arguments in [18], Thm. 19.7. The same situation holds for problems (P) involving arbitrary differential inclusions (1.2) with $x \in \mathbf{R}$; see [18], Rem. 19.2.

So the relaxation stability is an internal property of optimal control problems like (P) that is somehow inherent in a large class of continuous-time systems due to their "hidden convexity." It turns out that this property is actually *necessary and sufficient* for the *value convegence* of discrete apoproximations under *appropriate perturbations* of endpoint constraints.

To provide a precise formulation of this result we further assume that there is a minimizing sequence of feasible solutions to (P) belonging to the set W in (H1) and, in addition, one has:

(H2) The functions $\varphi_i,\ i = 0, 1, \ldots, q + r$, are continuous in $W \times W$ while the set Ω is closed.

2.2. Theorem. *If (P) is stable with respect to relaxation, then there exists a sequence of perturbations* $\{\gamma_{iN}, \eta_{iN}, \rho_N\} \downarrow 0$ *as* $N \to \infty$ *in* (2.3)–(2.5) *such that the value convergence* $(P_N) \to (P)$ *holds. Moreover, the relaxation stability of* (P) *is also a necessary condition for the value convergence of its discrete approximations under arbitrary perturbations of endpoint constraints.*

The proof of this theorem is based on the uniform approximation of an arbitrary trajectory for (1.2) by extended trajectories for (2.1) (this follows from Theorem 2.1) and certain relaxation arguments as in [18], Sec. 10. In this way we can provide effective estimates of the endpoint perturbations $\{\gamma_{iN}, \eta_{iN}, \rho_N\} \downarrow 0$ in Theorem 2.2 matching the discretization step h_N as $N \to \infty$. Similar results also hold for integral cost functionals like in (2.8).

Let us observe that Theorem 2.2 does not require any information about optimal solutions to (P). Thus it supports a numerical procedure to solve (P) using a proper reduction to finite dimensional problems of discrete approximations. Moreover, it is easy to derive from Theorem 2.2 by standard arguments that (extended) optimal trajectories $\bar{x}_N(\cdot)$ in (P_N) *uniformly* converge to an optimal solution $\bar{x}(\cdot)$ of (P) when the latter is stable with respect to relaxation.

Now let us consider another situation when an optimal solution $\bar{x}(\cdot)$ is *given* and the main goal is to obtain necessary optimality conditions for differential inclusions using discrete approximation as a *vehicle.* In this case we can include $\bar{x}(\cdot)$ in the approximation process and consider a modified sequence of discrete approximation problems $(\bar{P}_N)$ as follows: minimize

$$\varphi_0(x_N(a), x_N(b)) + \|x_N(a) - \bar{x}(a)\|^2 + \sum_{t \in T_N} \int_t^{t+h_N} \|\frac{x_N(t + h_N) - x_N(t)}{h_N} - \dot{\bar{x}}_N(\tau)\|^2 d\tau \quad (19)$$

subject to (2.2)–(2.5). Such a modification allows us to obtain a stronger convergence result that ensures not only the uniform convergence of optimal discrete trajectories to $\bar{x}(t)$ but also the *pointwise* convergence of their velocities a.e. in $[a, b]$. The latter fact is crucial to derive the refined Euler-Lagrange condition (1.7) for nonconvex differential inclusions; see Section 4.

2.3. Theorem. *Let* $\bar{x}(\cdot)$ *be an optimal solution to problem (P) that is stable with respect to relaxation. Assume that* (H1) *and* (H2) *hold in some neighborhood* W *of* $\bar{x}(\cdot)$. *Then there is a sequence of constraint perturbations* $\{\gamma_{iN}, \eta_{iN}, \rho_N\} \downarrow 0$ *such that:*

(i) *The discrete approximation problems* $(\bar{P}_N)$ *have optimal solutions for all large* N.

(ii) *For any sequence of extended optimal trajectories* $\{\bar{x}_N(\cdot)\}$ *in* $(\bar{P}_N)$ *one has*

$$\max_{t \in [a,b]} \|\bar{x}_N(t) - \bar{x}(t)\| \to 0 \ \text{ and}$$

$$\int_a^b \|\dot{\bar{x}}_N(t) - \dot{\bar{x}}(t)\|^2 dt \to 0 \text{ as } N \to \infty.$$

The proof of this theorem is based on the strong approximation result in Theorem 2.1. An analog of Theorem 2.3 holds also for integral functionals (2.8) where the integrand $g(x, v, t)$ is continuous in (x, v) and *measurable* in t; cf. [22], Thm. 3.3 and Rem. 3.4.

The convergence results established above allow us to make a bridge between dynamic optimization problems for discrete and differential inclusions. Discrete-time problems can be reduced, in turn, to finite dimensional problems of mathematical programming with *many geometric constraints* that may have empty interiority. The latter problems are definitely *nonsmooth* and require special tools of generalized differentiation we consider next.

3 Generalized Differentiation

This section is concerned with basic tools of generalized differentiation that are more appropriate for the main purposes and methods of this study. The results presented are mostly connected with the dual-space geometric approach in Mordukhovich [16–18] and recent developments in [21, 24]. We also refer the reader to Aubin and Frankowska [1], Clarke [5, 6], Ioffe [9, 10], Loewen [13], Rockafellar and Wets [31], and bibliographies therein for related and additional material.

Let Ω be a nonempty set in $\mathbf{R}^n$, and let

$$\Pi(x; \Omega) := \{\omega \in \text{cl}\,\Omega \;\text{ s.t. }\; \|x - \omega\| = \text{dist}(x; \Omega)\}$$

be the projector of x on Ω with the Euclidean distance function $\text{dist}(x, \Omega)$. The closed (maybe nonconvex) cone

$$N(\bar{x}; \Omega) := \text{Limsup}_{x \to \bar{x}}[\text{cone}(x - \Pi(x; \Omega))] \tag{20}$$

is called the *normal cone* to Ω at the point $\bar{x} \in \text{cl}\,\Omega$.

When Ω is a convex set, this normal cone coincides with the normal cone of convex analysis. In general, (3.1) admits the following representation

$$N(\bar{x}; \Omega) = \text{Limsup}_{x(\in cl\ \Omega) \to \bar{x}} \hat{N}(x; \Omega) \tag{21}$$

where the convex cone

$$\hat{N}(x|\Omega) := \{x^\star \in \mathbf{R}^n |\ \limsup_{x'(\in\Omega) \to x} \|x' - x\|^{-1} \langle x^\star, x' - x\rangle \le 0\} \tag{22}$$

coincides with the polar to the well-known (Bouligand) *contingent cone*. Moreover, the convex closure of (3.1) is always equal to the *Clarke normal cone*:

$$\tilde{N}(\bar{x}; \Omega) = \text{clco}\, N(\bar{x}; \Omega). \tag{23}$$

Despite its nonconvexity, the normal cone (3.1) enjoys many important calculus and related properties that allow us to use it for the analysis of nonsmooth problems. Some of these properties may worsen considerably when one takes the convex hull as in (3.4); see below. Note, in particular, that (3.1) is always *robust*:

$$N(\bar{x}; \Omega) = \text{Limsup}_{x \to \bar{x}} N(x; \Omega) \tag{24}$$

unlike (3.3) and (3.4) in general settings.

Next we consider a multifunction $F : X \rightrightarrows Y$ of the closed graph

$$\text{gph } F := \{(x,y) \in \mathbf{R}^n \times \mathbf{R}^m |\ y \in F(x)\}$$

and construct a derivative-like object for F using the normal cone (3.1) to its graph. The multifunction $D^\star F(\bar{x},\bar{y}) : \mathbf{R}^m \rightrightarrows \mathbf{R}^n$ defined by

$$D^\star F(\bar{x},\bar{y})(y^\star) := \{x^\star \in \mathbf{R}^n |\ (x^\star, -y^\star) \in N((\bar{x},\bar{y}); \text{gph } F)\} \tag{25}$$

is called the *coderivative* of F at the point $(\bar{x},\bar{y}) \in \text{gph } F$. The symbol $D^\star f(\bar{x})(y^\star)$ is used in (3.6) when $F = f$ is single-valued at $\bar{x}$.

According to this definition, the coderivative multifunction $D^\star F(\bar{x},\bar{y})(\cdot)$ is positive homogeneous with closed (maybe nonconvex or empty) values. It follows from (3.5) and (3.6) that the coderivative is robust with respect to perturbations of all the initial data. Moreover, due to (3.2) the coderivative (3.6) is a robust regularization of the dual construction

$$\hat{D}^\star F(\bar{x},\bar{y})(y^\star) := \{x^\star |\ \langle x^\star, v\rangle \le \langle y^\star, u\rangle \ \ \forall (u,v) \in \text{gph } DF(\bar{x},\bar{y})\}$$

to the *contingent derivative*

$$DF(\bar{x},\bar{y})(v) := \text{Lim}_{h\to v,\ \tau\downarrow 0}[F(\bar{x}+\tau h) - \bar{y}]/\tau$$

generated by the contingent cone to the graph of F. Let us observe that our basic construction (3.6) *cannot be dual* to any tangentially generated graphical derivative due to the *nonconvexity* of its values. Note that

$$D^\star f(\bar{x})(y^\star) = \{(\nabla f(\bar{x}))^\star\} \ \ \forall y^\star \in \mathbf{R}^m$$

when $f : \mathbf{R}^n \to \mathbf{R}^m$ is *strictly differentiable* at $\bar{x}$ with the Jacobian $\nabla f(\bar{x}) \in \mathbf{R}^{m\times n}$, i.e.,

$$\lim_{x,x'\bar{x}}[\frac{f(x) - f(x') - \nabla f(\bar{x})(x-x')}{\|x-x'\|}] = 0;$$

in particular, $f \in C^1$ around $\bar{x}$.

One of the most significant advantages of the coderivative (3.6) is its ability to provide *complete characterizations* of Lipschitzian and related properties of arbitrary closed-graph multifunctions. Such characterizations are particularly important for the method of discrete approximations (and other limiting procedures involving differential inclusions) since they imply the required convergence of adjoint arcs; see below. These questions and results are considered in detail in the paper [21]. Let us present the basic characterization of a general Lipschitz-like property of multifunctions introduced by Aubin: $F : \mathbf{R}^n \rightrightarrows \mathbf{R}^m$ is *pseudo-Lipschitzian* around $(\bar{x},\bar{y}) \in \text{gph } F$ with modulus $l \ge 0$ if there are neighborhoods W of $\bar{x}$ and V of $\bar{y}$ such that

$$F(x) \cap V \subset F(x') + l\|x - x'\|B \ \ \forall x, x' \in W.$$

This property of F turns out to be equivalent to certain fundamental properties of the inverse multifunction F^{-1} known as *metric regularity* and *openness at linear rate*; see [21] and references therein.

3.1. Theorem. *The following are equivalent:*

(i) *F is pseudo-Lipschitzian around $(\bar{x},\bar{y})$.*

(ii) $D^\star F(\bar{x},\bar{y})(0) = \{0\}$.

(iii) *There are neighborhoods* W *of* $\bar{x}$, V *of* $\bar{y}$ *and modulus* l *with*

$$\sup\{\|x^\star\| \text{ s.t. } x^\star \in D^\star F(x,y)(y^\star)\} \le l\|y^\star\| \;\; \forall x \in W,\; y \in F(x) \cap V,\; y^\star \in \mathbf{R}^m.$$

Moreover, the exact upper bound of moduli l *above is equal to the coderivative norm* $\|D^\star F(\bar{x},\bar{y})\|$.

Applying the null-criterion (ii) for the inverse F^{-1} and using the mentioned equivalence, one has a characterization of the metric regularity (openness) property of F around $(\bar{x},\bar{y})$ in the form of Ker $D^\star F(\bar{x},\bar{y}) = \{0\}$.

When F is locally bounded around $\bar{x}$, its classical Lipschitz behavior as in (H1) is equivalent to the pseudo-Lipschitzian property of F around $(\bar{x},\bar{y})$ for any $\bar{y} \in F(\bar{x})$. Therefore, Theorem 3.1 implies dual criteria for the classical Lipschitz continuity of multifunctions. In particular, this property is equivalent to the local *uniform boundedness* of the coderivative in (iii) holding for any $y \in F(x)$ and $x \in W$. The latter fact turns out to be *crucial* for limiting procedures involving Lipschitzian differential inclusions.

Despite the nonconvexity of values in (3.6), this coderivative possesses comprehensive calculus properties that are not generally available for its convex counterparts; see [24]. Let us present a coderivative *chain rule* that follows from Theorem 5.1 in [24].

For arbitrary closed-graph multifunctions $G : \mathbf{R}^m \Rightarrow \mathbf{R}^q$ and $F : \mathbf{R}^n \Rightarrow \mathbf{R}^m$ we consider their composition

$$(G \circ F)(x) := G(F(x)) = \bigcup_{y \in F(x)} G(y)$$

and express $D^\star(G \circ F)$ in terms of $D^\star G$ and $D^\star G$.

3.2. Theorem. *Let* $\bar{z} \in (G \circ F)(\bar{x})$ *and let the multifunction*

$$(x,z) \to F(x) \cap G^{-1}(z) = \{y \in F(x)|\; z \in G(y)\}$$

be locally bounded around $(\bar{x},\bar{z})$. *Then under the qualification condition*

$$D^\star G(\bar{y},\bar{z})(0) \cap \text{Ker } D^\star F(\bar{x},\bar{y}) = \{0\} \;\; \forall \bar{y} \in F(\bar{x}) \cap G^{-1}(\bar{z}) \tag{26}$$

one has the coderivative chain rule:

$$D^\star(G \circ F)(\bar{x},\bar{z}) \subset \bigcup_{\bar{y} \in F(\bar{x}) \cap G^{-1}(\bar{z})} D^\star F(\bar{x},\bar{y}) \circ D^\star G(\bar{y},\bar{z}).$$

The proof of this theorem and related calculus results for nonconvex constructions is based on an *extremal principle*; see [24]. It follows from Theorem 3.1 that the qualification condition (3.7) and therefore the chain rule in Theorem 3.2 automatically hold when, for all $\bar{y} \in F(\bar{x}) \cap G^{-1}(\bar{z})$, either G is pseudo-Lipschitzian around $(\bar{y},\bar{z})$ or F is metrically regular around $(\bar{x},\bar{y})$. The latter assumptions are fulfilled in many important situations and effectively support calculus rules for the coderivative (3.6) and associated subdifferential constructions in general settings; see [21, 24, 25] for more details.

Note that characterizations of Lipschitzian and related properties as in Theorem 3.1 and calculus rules as in Theorem 3.2 may fail if one considers another coderivative of $F : \mathbf{R}^n \Rightarrow \mathbf{R}^m$ generated by the Clarke normal cone

$$\tilde{D}^\star F(\bar{x},\bar{y})(y^\star) := \{x^\star \in \mathbf{R}^n|\; (x^\star, -y^\star) \in \tilde{N}(\bar{x},\bar{y}); \text{gph } F)\}. \tag{27}$$

One always has

$$\text{co}\, D^\star F(\bar{x},\bar{y})(y^\star) \subset \tilde{D}^\star F(\bar{x},\bar{y})(y^\star) \;\; \forall y^\star \in \mathbf{R}^m$$

where the inclusion is *proper* for a broad class of multifunctions whose graphs are nonsmooth *Lipschitzian manifolds* in the sense of Rockafellar [28], i.e., they are locally homeomorphic around $(\bar{x},\bar{y})$ to graphs of nonsmooth Lipschitz continuous functions. Besides locally Lipschitzian vector functions, this class includes maximal monotone operators and covers subdifferential mappings for a large core of functions important in variational analysis and optimization (e.g., convex, saddle, paraconvex, prox-regular functions, etc.; see [27, 28]).

It turns out that that for such multifunctions $\tilde{D}^\star$-analogs of the characterizing conditions (ii) and (iii) in Theorem 3.1 are *never fulfilled* outside of "strict smoothness." Moreover, the coderivatives (3.6) and (3.8) are different in dimensions; see [25], Sec. 3, for more details and related discussions. That is why the refined form of the Euler-Lagrange inclusion in (1.7) is distinctly sharper than Clarke's one involving (3.8).

The next result ([18], Thm. 3.1) shows that the Euler-Lagrange inclusion (1.7) and its discrete analog in terms of the basic coderivative (3.6) automatically imply the *maximum/minimum conditions* in problems with convex velocities. It what follows we use a conventional concept of *inner* (or lower) *semicontinuity* for multifunctions meaning that for every $\bar{y} \in F(\bar{x})$ and any sequence $x_k \to \bar{x}$ there is $y_k \in F(x_k)$ such that $y_k \to \bar{y}$ as $k \to \infty$.

3.3. Theorem. *Let F be convex-valued around $\bar{x}$ and inner semicontinuous at this point. Then for any $(\bar{y}, y^\star)$ with $D^\star F(\bar{x},\bar{y})(y^\star) \neq \emptyset$ one has*

$$\langle y^\star, \bar{y}\rangle = \min\{\langle y^\star, y\rangle |\; y \in F(\bar{x})\}.$$

Finally let us consider the basic subdifferential concept for extended-real-valued functions that corresponds to the coderivative (3.6) generated by the nonconvex normal cone (3.1). Let $\varphi : \mathbf{R}^n \to [-\infty, \infty]$ and let

$$E_\varphi(x) := \{\mu \in \mathbf{R}|\; \mu \geq \varphi(x)\}$$

be the epigraphical multifunction associated with φ. The set

$$\partial\varphi(\bar{x}) := D^\star E_\varphi(\bar{x},\varphi(\bar{x}))(1) = \{x^\star \in \mathbf{R}^n|\; (x^\star, -1) \in N((\bar{x},\varphi(\bar{x})); \text{epi}\,\varphi)\} \tag{28}$$

is called the *subdifferential]* of φ when $|\varphi(\bar{x})| < \infty$ (otherwise $\partial\varphi(\bar{x}) := \emptyset$).

When φ is lower semicontinuous (l.s.c.) around $\bar{x}$, the subdifferential (3.9) admits an equivalent analytical representation

$$\partial\varphi(\bar{x}) = \text{Limsup}_{x\to\bar{x},\; \varphi(x)\to\varphi(\bar{x})}\hat{\partial}\varphi(x)$$

where the construction

$$\hat{\partial}\varphi(x) := \{x^\star \in \mathbf{R}^n|\; \liminf_{x'\to x} \frac{\varphi(x') - \varphi(x) - \langle x^\star, x' - x\rangle}{\|x' - x\|} \geq 0\}$$

is well known in nonsmooth analysis (it is often called the *Fréchet subdifferential*) and coincides in fact with the subdifferential in the sense of *viscosity solutions.* If φ is locally Lipschitzian around $\bar{x}$, the convexification of (3.9) gives the Clarke generalized gradient.

When $f : \mathbf{R}^n \to \mathbf{R}^m$ is a Lipschitz continuous vector function, the coderivative of f can be represented in terms of the subdifferential (3.9) as follows

$$D^\star f(\bar{x})(y^\star) = \partial\langle y^\star, f\rangle(\bar{x}) \ \ \forall y^\star \in \mathbf{R}^m$$

where $\langle y^\star, f\rangle(x) := \langle y^\star, f(x)\rangle$ is the Lagrange scalarization of f. Moreover, the convexification of $D^\star f(\bar{x})$ is related to Clarke's generalized Jacobian $Jf(\bar{x})$ by

$$\text{co } D^\star f(\bar{x}(y^\star) = \{A^\star y^\star |\ A \in Jf(\bar{x})\} \ \ \forall y^\star \in \mathbf{R}^m.$$

It is easy to observe that

$$N(\bar{x}; \Omega) = \partial\delta(\bar{x}; \Omega) \ \ \forall \bar{x} \in \Omega$$

where $\delta(x; \Omega)$ is the (extended-real-valued) *indicator function* of Ω defined as 0 for $x \in \Omega$ and ∞ for $x \notin \Omega$. Furthermore, the normal cone (3.1) can be expressed as the subdifferential (3.9) of the (Lipschitz continuous) *distance function* $\text{dist}(x; \Omega)$ as follows:

$$N(\bar{x}; \Omega) = \text{cone}[\partial \text{dist}(\bar{x}; \Omega)] \ \ \forall \bar{x} \in \text{cl } \Omega.$$

Various calculus results involving the subdifferential (3.9) and the normal cone (3.1) can be easily deduced from the corresponding coderivative calculus mentioned above. Let us present two well-known rules used in the sequel:

$$N(\bar{x}; \Omega_1 \cap \Omega_2) \subset N(\bar{x}; \Omega_1) + N(\bar{x}; \Omega_2)$$

for any closed sets $\Omega_i,\ i = 1, 2$, with $N(\bar{x}; \Omega_1) \cap (-N(\bar{x}; \Omega_2)) = \{0\}$; and

$$\partial(\varphi_1 + \varphi_2)(\bar{x}) \subset \partial\varphi_1(\bar{x}) + \partial\varphi_2(\bar{x}) \tag{29}$$

if both functions $\varphi_i,\ i = 1, 2$, are l.s.c. and one of them is locally Lipschitzian around $\bar{x}$. We refer the reader to [6, 9, 13, 18, 24, 31] for more results and discussions in this direction.

4 Necessary Optimality Conditions

First let us consider a general finite dimensional problem of *nondifferentiable programming* in the following form:

$$\text{minimize } \ \varphi_0(x) \tag{30}$$

$$\text{subject to } \ x \in \Omega \subset \mathbf{R}^d \ \text{ and} \tag{31}$$

$$\varphi_i(x) \le 0 \ \text{ for } \ i = 1, 2, \ldots, q; \tag{32}$$

$$\varphi_i(x) = 0 \ \text{ for } \ i = q + 1, q + 2, \ldots, q + r; \tag{33}$$

$$x \in \Omega_j \ \text{ for } \ j = 1, 2, \ldots, s. \tag{34}$$

This problem contains functional constraints of inequality and equality type as well as many geometric constraints one of which (4.2) is singled out. Such a form allows us to reduce the basic discrete approximation problem (P_N) in Section 2 to the framework (4.1)–(4.2) where (4.5) corresponds to dynamic constraints.

Let us define the *essential Lagrangian*

$$L_\Omega(x, \lambda_0, \lambda_1, \ldots, \lambda_{q+r}) := \lambda_0 \varphi_0(x) + \ldots + \lambda_{q+r}\varphi_{q+r}(x) + \delta(x; \Omega) \tag{35}$$

in (4.1)–(4.5) and denote by $\partial_x L_\Omega(\cdot, \lambda_0, \ldots, \lambda_{q+r})$ the subdifferential (3.9) of (4.6) with respect to x. The next theorem corresponds to the *generalized Lagrange multiplier rule* in [18], Sec. 7, that was originally proved by using special (metric) approximations of (4.1)–(4.5) by a family of *unconstrained* optimization problems with *smooth* data; cf. [16–18].

4.1. Theorem. *Let $\bar{x}$ be an optimal solution to problem* (4.1)–(4.5) *where the functions φ_i, $i = 0, 1, \ldots, q + r$, are locally Lipschitzian and the sets Ω_j, $j = 0, 1, \ldots, s$, are closed around $\bar{x}$. Then there are multipliers $\lambda_0, \lambda_1, \ldots, \lambda_{q+r}$ and vectors*

$$x_j^\star \in N(\bar{x}; \Omega_j) \ \text{ for } \ j = 1, 2, \ldots, s,$$

not all zero, such that

$$\lambda_i \geq 0 \ \text{ for } \ i = 0, 1, \ldots, q, \tag{36}$$

$$\lambda_i \varphi_i(\bar{x}) = 0 \ \text{ for } \ i = 1, 2, \ldots, q, \ \text{ and}$$

$$-x_1^\star - x_2^\star - \cdots - x_s^\star \in \partial_x L_\Omega(\bar{x}, \lambda_0, \lambda_1, \ldots, \lambda_{q+r}).$$

Now reducing (P_N) to (4.1)–(4.5) for any fixed $N = 1, 2, \ldots$, we arrive at the following necessary optimality conditions in discrete approximations.

4.2. Theorem. *Let $\bar{x}_N(t)$, $t \in T_N \cup \{b\}$, be an optimal solution to (P_N) where the set Ω and the graph of $F(\cdot, t)$ are closed while the functions φ_i are Lipschitz continuous around $(\bar{x}(a), \bar{x}(b))$. Then there exist numbers λ_{iN} for $i = 0, 1, \ldots, q + r$ and an adjoint trajectory $p_N(t)$, $t \in T_N \cup \{b\}$, not all zero, such that one has the Euler-Lagrange inclusion*

$$\frac{p_N(t + h_N) - p_N(t)}{h_N} \in D_x^\star F(\bar{x}_N(t), \frac{\bar{x}(t + h_N) - \bar{x}_N(t), t}{h_N})(-p_N(t + h_N)) \ \text{ for } \ t \in T_N, \tag{37}$$

and the transversality inclusion

$$(p_N(a), -p_N(b)) \in \partial_x L_\Omega(\bar{x}_N(a), \bar{x}_N(b), \lambda_{0N}, \lambda_{1N}, \ldots, \lambda_{q+r\, N}) \tag{38}$$

supplemented by the sign and complementary slackness conditions

$$\lambda_{iN} \geq 0 \ \text{ for } \ i = 0, 1, \ldots, q, \tag{39}$$

$$\lambda_{iN}(\varphi_i(\bar{x}_N(a), \bar{x}_N(b)) - \gamma_{iN}) = 0 \ \text{ for } \ i = 1, 2, \ldots, q. \tag{40}$$

Let us observe that the necessary optimality conditions in Theorem 4.2 are obtained with *no* convexity and Lipschitzness assumptions on the multifunction F. These conditions follow directly from the generalized Lagrange multiplier rule in Theorem 4.1 and have an *Euler-Lagrange form* reflecting discrete dynamics.

Note that the transversality inclusion (4.9), expressed in terms of the nonconvex subdifferential (3.9) of the essential Lagrangian (4.6), strictly implies due to (3.10) more familiar "separated constraint" conditions

$$\begin{aligned}(p_N(a), -p_N(b)) &\in \partial(\sum_{i=0}^{q+r}\lambda_i\varphi_i)(\bar{x}(a), \bar{x}(b)) + N((\bar{x}(a), \bar{x}(b)); \Omega)\\ &\in \sum_{i=0}^{q}\lambda_i\partial\varphi_i(\bar{x}(a), \bar{x}(b)) + \sum_{i=q+1}^{q+r}\lambda_i\partial^0\varphi_i(\bar{x}(a), \bar{x}(b)) + N((\bar{x}(a), \bar{x}(b)); \Omega)\end{aligned}$$

where $\partial^0\varphi(\bar{x}) := \partial\varphi(\bar{x}) \cup [-\partial(-\varphi)(\bar{x})]$.

Let us consider the *adjoint inclusion* (4.8) that is an appropriate analog, in the general setting under consideration, of the classical Euler-Lagrange equation as well as the adjoint system in discrete optimal control. When F is *convex-valued* and inner semicontinuous in x around $\bar{x}(\cdot)$, the Euler-Lagrange inclusion (4.8) automatically implies, due to Theorem 3.3, the *discrete maximum condition*

$$\langle p_N(t+h_N), \frac{\bar{x}_N(t+h_N) - \bar{x}_N(t)}{h_N}\rangle = H(\bar{x}_N(t), p_N(t+h_N), t), \;\; t\imath n T_N \tag{41}$$

in terms of the Hamiltonian (1.9). In the case of standard optimal control problems for discrete-time systems like

$$x(t+h) = x(t) + hf(x(t), u(t), t), \;\; u(t) \in U(t), \tag{42}$$

the result of Theorem 4.2 coincides with the *discrete maximum principle* that is a discrete analog of the classical Pontryagin maximum principle for continuous-time systems (1.6); see [18] and references therein.

It is well known that the (exact) maximum condition (4.12) may be violated even for simple discrete-time systems with nonconvex velocity sets. This means that the discrete maximum principle, in contrast to its continuous-time counterpart, does not generally hold without convexity assumptions. As for systems like (1.6), they always have "hidden convexity" due to the time continuity (nonatomic measure). This is actually the main factor ensuring the Pontryagin maximum principle for (1.6) with no convexity assumed a priori.

On the other hand, discrete approximation systems, regarded as a *process* while $h \downarrow 0$, occupy an intermediate position between control systems with continuous and discrete (fixed h) time. It has been proved in [18, 19] that problems (P_N) with smooth dynamics (4.13) and endpoint constraints (2.3), (2.4) admit necessary optimality conditions in the form of *approximate maximum principle* where the exact maximum condition (4.12) is replaced by the approximate one

$$\langle p_N(t+h_N), \frac{\bar{x}_N(t+h_N) - \bar{x}_N(t)}{h_N}\rangle = H(\bar{x}_N(t), p_N(t+h_N), t) + \varepsilon(t, h_N) \tag{43}$$

with $\varepsilon(t,h) \to 0$ as $h \downarrow 0$ uniformly in t. This result was proved with *no convexity* assumptions on (4.13) under the *matching condition*

$$\eta_{iN}/h_N \to \infty \;\text{ as }\; N \to \infty \;\text{ for }\; i = q+1, \ldots, q+r$$

on admissible perturbations of equality type endpoint constraints. Examples show that the approximate maximum principle does not hold if the matching condition is violated.

The proof of the approximate maximum principle in [18, 19] essentially uses certain properties of smooth dynamics in (4.13) under which an approximate analog of "hidden convexity" is revealed. However, we strongly believe that (4.14) accomplishes necessary optimality conditions (4.8)–(4.11) in general nonconvex settings, i.e., the approximate maximum principle reflects intrinsic properties of discrete approximations as in (2.1)–(2.5) for nonconvex differential inclusions. A natural way to prove this general result consists of reducing discrete approximation problems for nonconvex differential inclusions to ones involving control systems with smooth dynamics where the approximate maximum principle has been already justified. For the case of continuous-time systems such a reduction was recently furnished in [33].

Considering discrete approximation problems $(\bar{P}_N)$ in Section 2, we can obtain necessary optimality conditions that are similar to the basic case of (P_N). Actually they follow from the conditions for (P_N) and distinguish from the latter only by terms vanishing as $N \to \infty$; cf. [22].

Finally let us go back to the original optimal control problem (P) for nonconvex differential inclusions. The next theorem summarizes the strongest necessary optimality conditions obtained for this problem using discrete approximations as well as other techniques; see the sketch of the proof.

4.3. Theorem. *Let $\bar{x}(\cdot)$ be an optimal solution to problem (P). Assume that F is locally Lipschitzian in x with a summable modulus, measurable in t and bounded by a summable function around $\bar{x}(t)$ a.e. in $[a,b]$, and that φ_i, $i = 0, 1, \ldots, q+r$, are locally Lipschitzian around $(\bar{x}(a), x(\bar{b})$ while Ω is closed. Then there are numbers $\lambda_0, \lambda_1, \ldots, \lambda_{q+r}$ satisfying (4.7) and an absolutely continuous function $p : [a,b] \to \mathbf{R}^n$, not all zero, such that one has:*
the Euler-Lagrange inclusion

$$\dot{p}(t) \in \text{co } D_x^\star F(\bar{x}(t), \dot{\bar{x}}(t), t)(-p(t)) \text{ a.e. } t \in [a,b], \tag{44}$$

the transversality condition

$$(p(a), -p(b)) \in \partial_x L_\Omega(\bar{x}(a), \bar{x}(b), \lambda_0, \lambda_1, \ldots, \lambda_{q+r}) \tag{45}$$

accomplished by the complementary slackness

$$\lambda_i \varphi_i(\bar{x}(a), \bar{x}(b))) = 0 \text{ for } i = 1, 2, \ldots, q+r, \tag{46}$$

and the maximum condition

$$\langle p(t), \dot{\bar{x}}(t) \rangle = H(\bar{x}(t), p(t), t) \text{ a.e. } t \in [a,b]. \tag{47}$$

When, in addition, the sets $F(x,t)$ are convex around $\bar{x}(\cdot)$, the Hamiltonian inclusion holds:

$$\dot{p}(t) \in \text{co } \{w| \; (-w, \dot{\bar{x}}(t)) \in \partial H(\bar{x}(t), p(t), t)\} \text{ a.e. } t \in [a,b]. \tag{48}$$

Sketch of the Proof. Some results of the theorem can be proved directly by the method of discrete approximations taking the limit in necessary optimality conditions for discrete-time systems. Indeed, due to Theorem 2.3 we approximate the given optimal solution $\bar{x}(\cdot)$ to (P) by optimal solutions $\bar{x}_N(\cdot)$ to discrete-time problems such that $\dot{\bar{x}}_N(t) \to \dot{\bar{x}}(t)$ *pointwisely* a.e. in $[a,b]$. Then passing to the limit in necessary optimality conditions for discrete approximations (Theorem 4.2) and using the tools of nonsmooth analysis presented in Section 3, we arrive at the Euler-Lagrange inclusion (4.15) and the transversality condition (4.16) accomplished by (4.17). Note that the convex hull of the coderivative appears in (4.15) in contrast to (4.8) since we just have the L^1-weak convergence of the adjoint derivatives $\dot{p}_N(t) \to \dot{p}(t)$ (by Theorem 3.1) and should use Mazur's convexification theorem to get the required pointwise convergence a.e. in $[a, b]$. This procedure works also for the case of Bolza-type cost functionals with integrands merely *measurable* in t; see [22, 23] for more details.

Let us observe that the application of Theorem 2.3 requires the relaxation stability of the original problem as well as a.e. continuity in t of F in (H1). However, both of these assumptions can be dropped by reducing (P) to an unconstrained Bolza problem (as in [22], Sec. 7) that is always stable with respect to relaxation due to Bogoljubov's theorem (2.8). In this way we obtain necessary optimality conditions (4.15)–(4.17) under the mild assumptions made; cf. [22, 23].

By Theorem 3.3 the maximum condition (4.18) follows directly from the Euler-Lagrange inclusion (4.15) when F is convex-valued around $\bar{x}(\cdot)$. In the nonconvex setting it can be obtained together with (4.15)–(4.17) passing to the limit as $N \to \infty$ in the *approximate* discrete maximum principle (4.14), (4.8)–(4.11). The latter result, however, has not been justified yet in full generality; see the discussion above. Available proofs (cf. [10, 23, 33]) of the maximum condition (4.18) supplementing (4.8)–(4.11) for nonconvex differential inclusions are based on reducing (P) to an unconstrained problem of Bolza where the Euler and Weierstrass conditions (with the same adjoint arc) have been recently obtained by Ioffe and Rockafellar [11].

When F is *convex-valued* around $\bar{x}(\cdot)$, the Hamiltonian inclusion (4.19) is *equivalent* to the Euler-Lagrange inclusion (4.15) under the assumptions made above. This follows from equivalence results of Rockafellar [30] for extended-real-valued integrands with certain epi-continuity properties. Ioffe [10] recently found a general framework of unbounded differential inclusions satisfying mild Lipschitz-type requirements under which the Euler-Lagrange condition (4.15) holds without convexity and always implies the Hamiltonian condition (4.18) when $F(x, t)$ are convex. This emphasizes a *pivotal role of the Euler-Lagrange condition* in variational analysis.

In conclusion we observe that the necessary optimality conditions in Theorem 4.3 are expressed in terms of *robust* derivative-like constructions for nonsmooth objects being stable with respect to perturbations of the initial data. Therefore, one can freely use these conditions for sensitivity analysis and other applications involving perturbation/approximation procedures.

References

1. J.-P. Aubin, H. Frankowska: *Set-Valued Analysis.* Birkhäuser, Boston 1990.

2. N.N. Bogoljubov: *Sur quelques methodes nouvelles dans le calculus des variations.* Ann. Math. Pure Appl. 7, 1930, 249–271.

3. F.H. Clarke: *Optimal solutions to differential inclusions.* J. Optim. Theory Appl. 19, 1976, 460–478.

4. F.H. Clarke: *Necessary conditions for a general control problem.* In: "Calculus of Variations and Control Theory" (D. Russel, ed.), Academic Press, N.Y., 1976, 257–278.

5. F.H. Clarke: *Optimization and Nonsmooth Analysis.* Wiley-Interscience, N.Y. 1983.

6. F.H. Clarke: *Methods of Dynamic and Nonsmooth Optimization.* SIAM Publications, Philadelphia, PA, 1989.

7. A.D. Dontchev, F. Lempio: *Difference methods for differential inclusions: a survey.* SIAM Rev. 34, 1992, 263–294.

8. H. Frankowska: *Contingent cones to reachable sets of control systems.* SIAM J. Contr. Optim. 27, 1989, 170–198.

9. A.D. Ioffe: *Approximate subdifferentials and applications, I: the finite dimensional theory.* Trans. Amer. Math. Soc. 281, 1984, 289–316.

10. A.D. Ioffe: *Euler-Lagrange and Hamiltonian formalisms in dynamic optimization.* Trans. Amer. Math. Soc., to appear.

11. A.D. Ioffe, R.T. Rockafellar: *The Euler and Weierstrass conditions for nonsmooth variational problems.* Calculus of Variations and Partial Differential Equations 6, 1996, 59–87.

12. B. Kaśkosz and S. Lojasiewicz, Jr.: *Lagrange-type extremal trajectories in differential inclusions.* Systems Control Lett. 19, 1992, 241–247.

13. P.D. Loewen: *Optimal Control via Nonsmooth Analysis.* American Mathematical Society, Providence, RI, 1993.

14. P.D. Loewen, R.T. Rockafellar: *Optimal control of unbounded differential inclusions.* SIAM J. Optim. Control 32, 1994, 442–470.

15. P.D. Loewen, R.T. Rockafellar: *New necessary conditions for the generalized problem of Bolza.* SIAM J. Control Optim. 34, 1996, 1496–1511.

16. B.S. Mordukhovich: *Maximum principle in problems of time optimal control with nonsmooth constraints.* J. Appl. Math. Mech. 40, 1976, 960–969.

17. B.S. Mordukhovich: *Metric approximations and necessary optimality conditions for general classes of nonsmooth extremal problems.* Soviet Math. Dokl. 22, 1980, 526–530.

18. B.S. Mordukhovich: *Approximation Methods in Problems of Optimization and Control.* Nauka, Moscow 1988.

19. B.S. Mordukhovich: *Approximate maximum principle for finite difference control systems.* U.S.S.R. Comput. Maths. Math. Phys. 28, 1988, 106–114.

20. B.S. Mordukhovich: *On variational analysis of differential inclusions.* In: "Optimization and Nonlinear Analysis" (A.D. Ioffe et al., eds.), Longman, Harlow, 1992, 199–213.

21. B.S. Mordukhovich: *Complete characterization of openness, metric regularity, and Lipschitzian properties of multifunctions.* Trans. Amer. Math. Soc. 340, 1993, 1–35.

22. B.S. Mordukhovich: *Discrete approximations and refined Euler-Lagrange conditions for nonconvex differential inclusions.* SIAM J. Control Optim. 33, 1995, 882–915.

23. B.S. Mordukhovich: *Necessary optimality and controllability conditions for nonsmooth control systems.* Proceedings of the 33rd Conference of Decision and Control, Lake Buena Vista, FL, 1994, 3992–3997.

24. B.S. Mordukhovich: *Generalized differential calculus for nonsmooth and set-valued mappings.* J. Math. Anal. Appl. 183, 1994, 250–288.

25. B.S. Mordukhovich: *Stability theory for parametric generalized equations and variational inequalities via nonsmooth analysis.* Trans. Amer. Math. Soc. 343, 1994, 609–658.

26. B.S. Mordukhovich: *Optimization and finite difference approximations of differential inclusions with free time.* In: "Nonsmooth Analysis and Geometric Methods in Deterministic Optimal Control" (B.S. Mordukhovich and H.J. Sussmann, eds.), Springer-Verlag, N.Y., 1996, 153–202.

27. R.A. Poliquin, R.T. Rockafellar: *Prox-regular functions in variational analysis.* Trans. Amer. Math. Soc. 348, 1996, 1805–1838.

28. R.T. Rockafellar: *Maximal monotone relations and the second derivatives of nonsmooth functions.* Ann. Inst. H. Poincaré. Anal. Non Linéair 2, 1985, 167–184.

29. R.T. Rockafellar: *Dualization of subgradient conditions for optimality.* Nonlinear Anal. Theory Methods Appl. 20, 1993, 627–646.

30. R.T. Rockafellar: *Equivalent subgradient versions of Hamiltonian and Euler-Lagrange equations in variational analysis.* SIAM J. Control Optim. 34, 1996, 1300–1314.

31. R.T. Rockafellar, R.J-B Wets: *Variational Analysis.* Springer-Verlag, N.Y., 1997.

32. G.V. Smirnov: *Discrete approximations and optimal solutions of differential inclusions.* Cybernetics 27, 1991, 101–107.

33. R.B. Vinter, H. Zheng: *The extended Euler-Lagrange condition for nonconvex variational problems.* SIAM J. Control Optim. 35, 1997, 56–77.

34. J. Warga: *Controllability, extremality, and abnormality in nonsmooth optimal control.* J. Optim. Theory Appl. 41, 1983, 239–260.

35. Q.J. Zhu: *Necessary optimality conditions for nonconvex differential inclusions with endpoint constraints.* J. Diff. Equat. 124, 1996, 186—204.